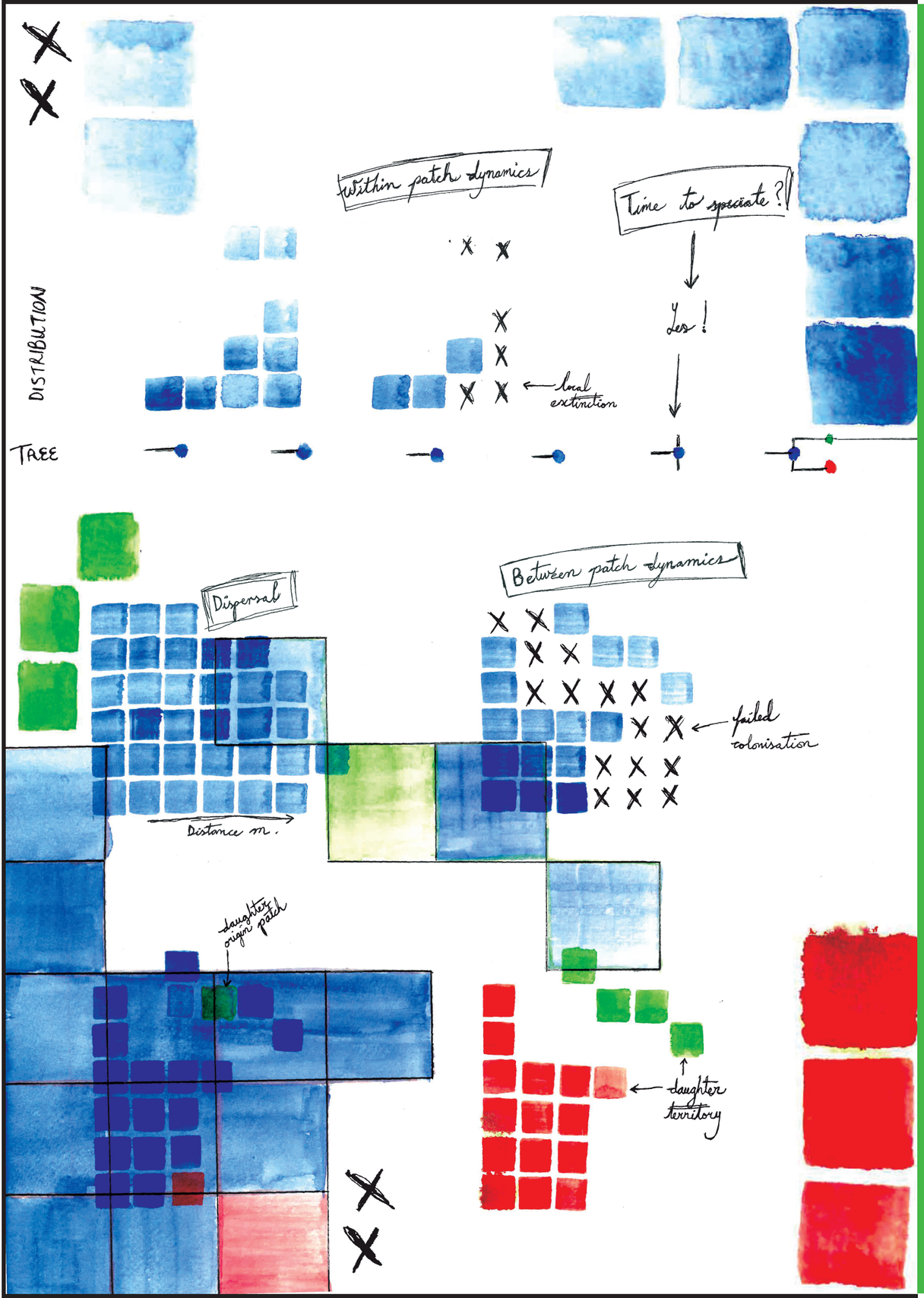
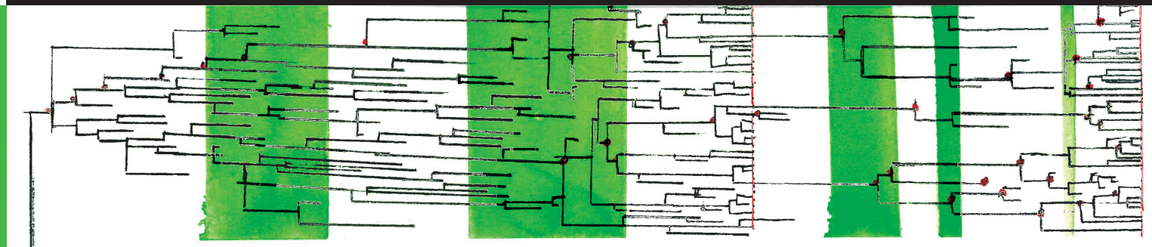


Ph.D.

2019

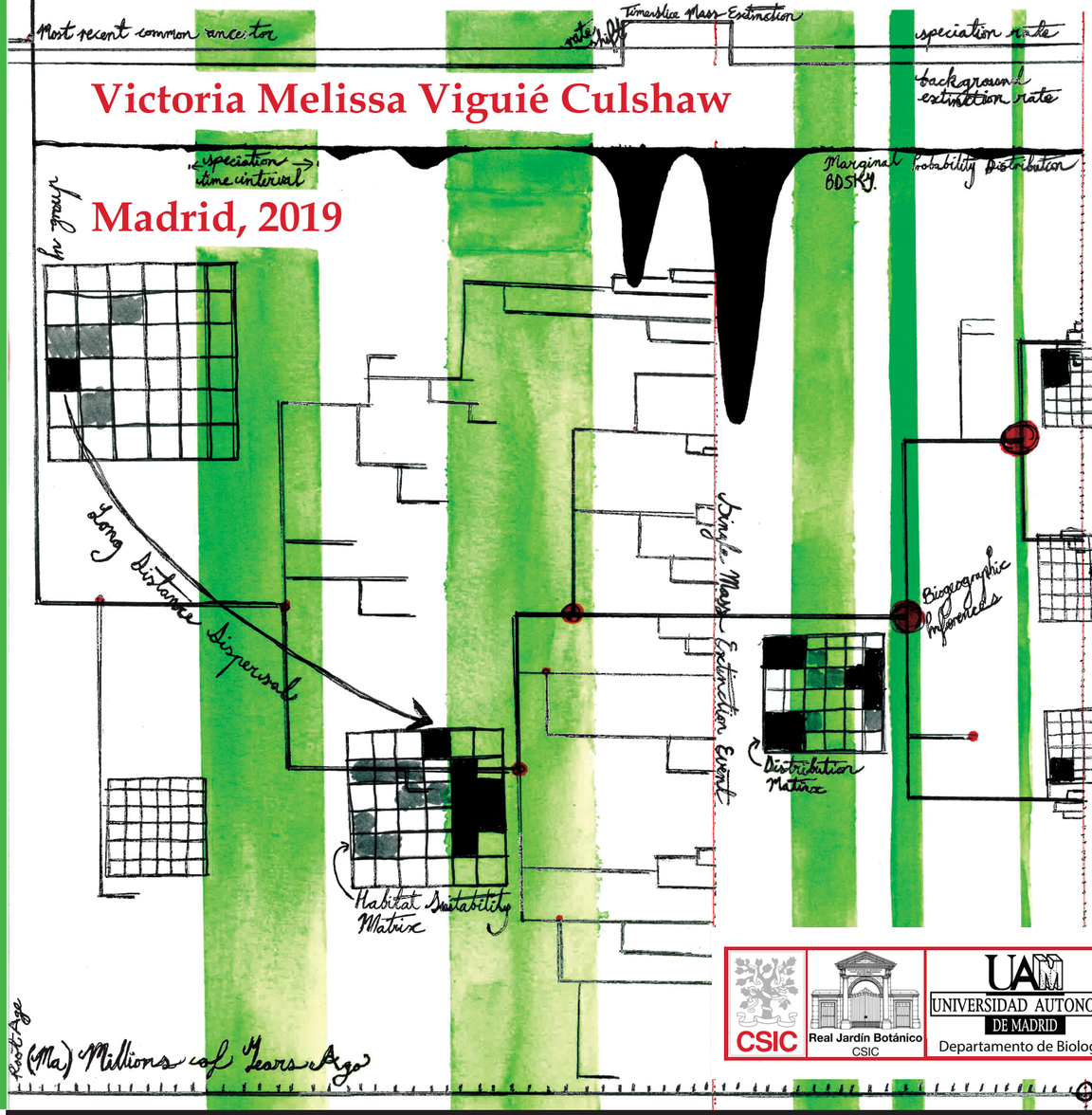
CULSHAW, V. M. V. C.



DEVELOPING NEW TOOLS TO ADDRESS THE IMPACT OF CLIMATE CHANGE ON THE EVOLUTIONARY AND DISTRIBUTIONAL HISTORY IN PLANT LINEAGES

Victoria Melissa Viguié Culshaw

Madrid, 2019



CSIC Real Jardín Botánico CSIC
UAM UNIVERSIDAD AUTONOMA DE MADRID
Departamento de Biología

DEVELOPING NEW TOOLS TO ADDRESS THE IMPACT OF CLIMATE CHANGE ON THE EVOLUTIONARY AND DISTRIBUTIONAL HISTORY IN PLANT LINEAGES

TESIS DOCTORAL

VICTORIA MELISSA VIGUIÉ CULSHAW

MADRID, 2019

DIRECTORA

ISABEL SANMARTÍN BASTIDA

TUTOR

JUAN CARLOS MORENO SAIZ

UAM
UNIVERSIDAD AUTONOMA
DE MADRID

Departamento de Biología

Este trabajo ha sido realizado en el Real Jardín Botánico de Madrid, CSIC, bajo la dirección de la doctora Desentrañando Isabel Sanmartín Bastida, científica titular del Departamento de Conservación y Biodiversidad. Además, parte de los resultados fueron obtenidos en las estancias realizadas en el grupo de Computational Evolution (cEvo), el Departamento de Ciencia de Biosistemas e Ingeniería (D-BSSE) en la Escuela Politécnica Federal de Zúrich (Eidgenössische Technische Hochschule, ETH), Zurich (Basel, Suiza) bajo la dirección de la doctora Tanja Stadler, en le Centro de I+D receptor: Laboratório de Ecologia Teórica e Síntese (LETS), ICB, Campus II/ Universidade Federal de Goiás (UFG, Goiânia, Goiás, Brasil) bajo la dirección del doctor Thiago F. L. V. B. Rangel. La realización de la tesis ha sido posible gracias a una ayuda predoctoral de Formación del Personal Investigador (FPI, BES-2013-065389) concedida a Victoria Melissa Viguié Culshaw por el Ministerio de Ciencia e Innovación (ahora Ministerio de Economía y Competitividad) y a los proyectos i) desentrañando el papel de la extinción asociada a cambios climáticos en el origen de la flora Africana (AFFLORA, proyecto CGL2012-40129-C02-01); y ii) desarrollo y aplicación de herramientas de secuenciación masiva e inferencia Bayesiana para detectar la huella filogenética de la extinción a diferentes escalas evolutivas (BAYESNEXT, CGL2015-67849-P), concedidos a Isabel Sanmartín por la misma entidad.

CSIC

Real Jardín Botánico
CSIC

UAM
UNIVERSIDAD AUTONOMA
DE MADRID

ETH zürich

cEvo

UFG
UNIVERSIDADE
FEDERAL DE GOIÁS

To Poppy cat,

for all the shoulder coories

and for wandering all over my laptop,

messing up my typing,

while looking adorable

and cute,

so cute.

TABLE OF CONTENTS

TABLE OF CONTENTS

AGRADECIMIENTOS

AGRADECIMIENTOS

Dear Isabel,

With the greatest of pride, I present to you my Ph.D. thesis entitled “Developing new tools to address the impact of climate change on the evolutionary and distributional history in plant lineages”. Thank you for taking a chance on me and welcoming me into your Rand Flora research group. Thank you for letting me tap into your knowledge, for all of your hard work, and for embracing all of my ideas and helping me realise the good ideas and bin the terrible ones. Our meetings were always epically long! By the end of them we had achieved a lot, but we were completely brain dead to the point that the healing affects of a simple cup of tea would not be great enough. The years we have spent together have had their fair share of rough and euphoric moments, and there were times where I did not think that I would finish, but here I am now, at the end of this Ph.D. program, I would say that for sure I am all the richer from this invaluable experience.

As with all great bodies of work, I cannot claim this work as solely mine, but instead it belongs to a collection of people who have believed in the work, in me, or both! So I wish to extend my thanks to them.

Thank you to my tutor Juan Carlos Moreno in el Universidad Autónoma de Madrid, for all of his help in my understanding of the Ph.D. program. Thank you for having sugar free chocolate biscuits available.

Thank you everyone of the Rand Flora team! I would like to make a specific mention to Alberto Herrero, Mario Mairal, Sanna Olson, Lisa Pokorny, Rikki Riina, and Tamara Villaverde. Thank you Alberto for all of your hard work and working with me in the laboratory. You made it so much fun and the work felt easy. We will forever be “¡¡equipo de sueño!!” and our secret hand shake will forever be an elbow high five. Thank you Mario for taking me under your wing when I began my thesis. Thank you for being one of my main proof readers. Thank you for introducing me to the Spanish culture and countryside, for teaching me how to collect, preserve and record plants for the herbarium collection –the memory of meeting the German hippy when we were collecting in Tenerife will always remain with me. I have learned a lot through you about plants and animals, and I saw a bear!!! Thank you for your dedication and faith in me to learn Spanish. Thank you Sanna for teaching and training me in the ways of the laboratory and teaching me how to edit sequences. Thank

you for showing me how to organise my work –the lab book and colour coded excel sheet certainly came in handy when I was confused, brain dead and cross-eyed from the hours spent editing sequences. Thank you Lisa for introducing me to and training me in the ways of the laboratory. I was pretty nervous in the beginning. I felt like a fish out of water there, you helped make me feel welcome. Thank you Riki for teaching me how to work in the laboratory, and to understand how to search through online herbarium databases and the RJB's own herbarium. I and my stomach thank you for all the slices of cake and delicious meals that you have presented to me over the years. Thank you Tamera for letting me speak to you about my second chapter, and for all of your advice whenever I needed it. Your motivation and enthusiasm helped propel me forward and finish. Thank you for sharing stories of your own academic experience. They were invaluable to me.

Thank you everyone in el Real Jardín Botánico, CSIC, Madrid. From the people working the front desk and the guards at the gate who feed the ducks, to all of the students. I have always felt welcome. It is nice to know that I will always be part of the garden as my voice is the English voice of the automated voice response system on the phone. I would like to thank in particular Iván Garcia and Gonzalo Nieto. Thank you Ivan for all of the smoking breaks, where you helped me laugh at my Ph.D. stories of misfortune and ridiculousness, and where we gossiped about news of the garden. Thank you for your support and advice throughout the years, for our many cañas, and for being the other member of "El Real Jardín Botánico Grupo del Cine" –maybe one day we will finally welcome a third member… I will forever be glad of the day I stole your PCR gel, for this day and a "sorry chocolate muffin" began a very nice friendship. Thank you Gonzalo for all of the advice that you have given me throughout the years and for having faith in me. Looks like I managed to keep up our end of the deal that you forced upon me in Montpellier –you pay for our dinner, I finish my thesis.

Thank you to all of my friends that I have made in Spain. In particular, I would like to thank Natalia Pozhilova and Giovanni Bernardi. Thank you Natalia for becoming like my sister, celebrating my successes, helping me in my lows. Thank you for taking me with you to visit your family in Russia when I was in need of a break and a different scene from my thesis, that was a most excellent trip. Thank you Gio for providing me with fresh perspective, for all of the beers in the underground bar and the bouldering in el Escorial.

Thank you to the cEvo team in D-BSSE, ETH, Basel, Switzerland for warmly welcoming me to your laboratory. In particular, I would like to thank Tanja Stadler, Veronika Boskova and Denise Kühnert. Thank you Tanja for helping me design the first chapter of this thesis, and for all of your advice, support and hard work during and after my stay until the chapter was successfully published. One memory that will stick into my mind is the time we had a meeting, me cycling beside you as you

ran. Thank you Veronika and Denise for all of your help in coding the BDKY model. Thank you everyone for introducing me to lunchtime sushi making Tuesdays, never before had I made my own sushi rice. I look forward to seeing you in future conferences, where we can share a beer and pizza together in the evenings.

Thank you to Joaquín Hortal for all of your encouragement for the third chapter and for giving me your invaluable contact in Brasil. The visit I made there was amazing. Thank you to the LETS team, Universidade Federal de Goiás, Brasil for welcoming me into your laboratory, I really enjoyed my stay. In particular, I would like to thank Thiago Rangel and Diego Llusia. Thank you Thiago for all of your advice and allowing me to tap into your knowledge whenever I was uncertain. Thank you Diego for introducing me to Latin dancing, and for our two week field trip to search for tree frogs and to install recorders to record their songs, it inspired me to always make sure that my thesis contains field work. We will always have "our Johnny" Cash –"*Johnny, Johnny!"*– to remember navigating the Brasilian roads during the trip.

Thank you to my German friends: Selasie Prince Segbedzi, Gjergj Bodari, Nilka Gordon, Genevieve Cyrs, Delini Anklam, Krischan Klein and Johannes Weisser. Your faith and pride in me and your support in the form of conversations, beer, food, and taking me outside to feel the sun on my skin and escape from my computer for a few hours has been wonderful. Thank you to Jakob Boeckh for the creation of this thesis's beautiful cover page. Now my thesis looks enticing, adorned in its finery.

Thank you to everyone in Scotland. In particular, I would like to thank Andrew Hendry, for it is he who helped me with the first ever version of Chapter 3, an in depth scribble after Peking duck, beer and a film. Thank you also to Ali McDougal for his proof reading and his supporting ear.

Thank you to my family. To my dad for giving me the dream of holding a Ph.D. title. To my mum for excitedly telling all of her friends and work colleges that "my daughter is a Ph.D. student", her parcels of Scottish food, and for her artistic insight to create many sexy figures for my chapters. My brothers Martin and Geoff for all of the time they have spent with me on the phone and in person listening to me and encouraging me. My sister-in-law Katrina for fiercely pep talking me whenever I lost courage or was tired, and equally as fiercely celebrating with me for any success I mentioned to her.

Thank you to my pen pal M. Le Paul Bastide, your words, time, unconditional support and perspective throughout these years have been amazing. Standing beside you in the train station after

the MCEB meet in Porquerolles was a happy and fortunate event to have occurred. I am glad to have learned, through you, of the ballet of water droplets making their way down a train window.

Thank you Jonas Möller for all of your time, advice and support that you have given me. Thank you for your whispering kisses, the many dishes of cheesy pasta, and of course German Pokémon.

Thank you to that person who first thought, all that time ago, a cat would make an excellent companion. I am forever in your debt.

Best Wishes,

Victoria M.V. Culshaw

Hamburg, Germany

November 2019

RESUMEN EN CASTELLANO

GENERAL INTRODUCTION

AIMS, HYPOTHESES AND OBJECTIVES

RESUMEN EN CASTELLANO

El impacto del meteorito “asesino de dinosaurios” Chicxulub supone el pistoletazo de salida del Cenozoico, la era que abarca los últimos 65 millones de años de la historia de la Tierra. Al igual que otras épocas del Fanerozoico, el Cenozoico se caracteriza por dramáticos cambios climáticos a distintas escalas espaciales, entre los que el Mioceno Superior destaca como un laboratorio climático para examinar el impacto de futuros escenarios de cambio climático: en este período, hubo incrementos rápidos de temperatura y descensos bruscos, y se dispone relativamente de abundantes datos ambientales. Estos cambios ambientales están relativamente bien documentados y proporcionan un marco ideal para entender las relaciones entre el clima, el ciclo del carbono, y la historia evolutiva y biogeográfica de las especies.

La tasa de extinción “de fondo” –el porcentaje de especies que se extinguen respecto a las que se originan– se ha estimado varios órdenes de magnitud más alta en el Antropoceno, la época geológica propuesta por la comunidad científica para suceder o reemplazar al Holoceno y definida como el período del Cuaternario afectado por la actividad humana, respecto a las tasas que se estiman para el registro fósil. Esto ha llevado a algunos científicos a hablar de la “sexta extinción masiva”, que provocará la desaparición de un porcentaje significativo de biodiversidad en un corto período de tiempo. Las extinciones masivas están ligadas a cambios climáticos intensos y rápidos o a eventos geológicos catastróficos (el K/T), y a menudo implican un cambio o reordenamiento en el ecosistema, en el que grupos dominantes antes del evento son reemplazados por otros. Si la actual tasa de extinción continúa a este ritmo, su efecto sería comparable en magnitud a los cinco grandes eventos previos de extinción masiva, en un tiempo no superior a 300 años. La principal consecuencia de la elevada tasa de extinción del Antropoceno sería una pérdida desproporcionada de historia evolutiva, entendida como millones de años de evolución de rasgos funcionales y morfológicos. Sin embargo, es complicado distinguir si esta pérdida se debe a la acción humana o a los ciclos naturales de expansión y declive de la tasas de diversificación típicos en la historia de nuestro planeta.

Es especialmente crítico mejorar nuestro conocimiento sobre el proceso y consecuencias de la extinción. Habría que conocer si las actuales especies amenazadas están abocadas a extinguirse o si por el contrario están experimentando un ciclo natural de diversificación lenta; si la tasa actual de extinción de fondo continuará aumentando en el futuro; y si las tasas de extinción estimadas en las especies mejor estudiadas pueden extrapolarse a otras especies. Este conocimiento nos permitiría una

definición más precisa de lo que es una especie amenazada, incluso si la extinción global no se produce hasta dentro de 10.000 años, y como resultado de esto, una mejor capacidad de decisión a la hora de determinar las prioridades en conservación.

El actual interés en el cambio climático ligado a la actividad humana, junto su impacto directo en la biodiversidad, ha despertado el interés por reconstruir la respuesta evolutiva de las especies a eventos de cambio climático en el pasado geológico. Para ello se han estudiado procesos de adaptación (incluyendo especiación), persistencia geográfica, cambios en el rango de distribución (migración, expansión, contracción), y extinción, bien local, global o masiva. Tradicionalmente, se han utilizado tres enfoques distintos para reconstruir la respuesta evolutiva y biogeográfica de los linajes al cambio climático: i. modelos macroevolutivos de especiación-extinción, que utilizan una filogenia con dimensión temporal para discriminar entre adaptación (radiación) y extinción global; ii. inferencia biogeográfica, basados en filogenias datadas asociadas a datos de distribución de las especies, que permiten inferir los rangos ancestrales de los linajes y eventos de migración histórica; iii. modelos de nicho climático, en los que la tolerancia climática de una especie, basada en registros de sus localidades de distribución, son proyectadas sobre escenarios climáticos pasados y futuros, para explorar la existencia de condiciones similares en distintas áreas o períodos de tiempo. Estos enfoques representan, respectivamente, las respuestas de las especies al cambio climático desde un punto de vista filogenético, biogeográfico, y ecológico. Existe, sin embargo, un cuarto enfoque, un enfoque integrador que pretende unir las disciplinas de la ecología, filogenia y biogeografía, (complementándose entre ellas para así superar sus limitaciones), con el objetivo común de desentrañar la señal evolutiva y biogeográfica del cambio climático en las especies.

En la línea de este enfoque integrador, esta tesis pretende desarrollar nuevas herramientas analíticas que combinan los modelos macroevolutivos, la inferencia biogeográfica, y la modelización del nicho climático para entender la respuesta de linajes de plantas al cambio ambiental acelerado, así como la creación de nuevos modelos computacionales de simulación en macroecología que incorporen la historia evolutiva de las especies. Estas nuevas herramientas se han aplicado a dos casos de estudio: la evolución del grupo de las coníferas (Gimnosperma, Coniferae), con distribución cosmopolita (Capítulo 1), y el género africano *Camptoloma*, que presenta una distribución continental disyunta conocida como Rand Flora (Capítulos 2 y 3).

El *Capítulo 1* explora el potencial de modelos macroevolutivos episódicos (BDSKY) para detectar y estimar eventos de extinción masiva (MEE) en filogenias moleculares que no contienen información del registro fósil. Al contrario que en otros métodos que modelizan los MEEs como un pulso instantáneo en el tiempo, BDSKY modeliza los MEEs como períodos de tiempo con una corta

duración en tiempo geológico, en los que la tasa de extinción supera a la tasa de especiación, seguidos por un período de recuperación con diversificación positiva. Se demuestra que este tipo de modelos permiten detectar la señal evolutiva del MEE en la filogenia, pero que la precisión y exactitud de la estima depende en gran medida del tamaño del árbol y la intensidad del evento. La comparación con otros métodos de "pulso único" como CoMET, demuestran que BDSKY tiene una frecuencia de error Tipo II muy parecida a este método y que apenas exhibe falsos positivos (error Tipo I). Sin embargo, BDSKY presenta menor poder estadístico para detectar MEEs que CoMET con filogenias pequeñas con pocos terminales, aunque muestra mayor exactitud a la hora de estimar las tasas de especiación y de extinción de fondo en filogenias de especies actuales.

El *Capítulo 2* propone un nuevo enfoque integrador para combinar la inferencia biogeográfica, dentro de un marco estadístico bayesiano, con modelos de nicho climático (ENM) proyectados espacialmente en el tiempo geológico. Al contrario que otras propuestas para integrar la información del análisis biogeográfico con ENMs, este enfoque no requiere de abundante información fósil o biogeográfica, si no que combina las fortalezas de cada uno de estos dos enfoques como evidencias independientes y complementarias para inferir la historia evolutiva de linajes relictos, con pocas especies y rangos de distribución reducidos como consecuencia del cambio climático. Este enfoque se aplica, utilizando como caso de estudio un pequeño género de tres especies –*Camptoloma* (Scrophulariaceae)– distribuido en las Islas Canarias y en el este y sur de África, en lo que se ha denominado el patrón biogeográfico Rand Flora. Combinando el uso de inferencia biogeográfica bayesiana, filogenia molecular basada en marcadores plastidiales y nucleares, y proyecciones ENM sobre capas paleoclimáticas, se reconstruye la evolución de este género como el resultado de eventos de dispersión y extinción climática ligadas a fenómenos de aridificación geológica (hipótesis de los "refugios climáticos"). Asimismo, se propone un nuevo enfoque para utilizar la evidencia evolutiva proporcionada por el análisis biogeográfico para seleccionar el valor de umbral que permita transformar probabilidades de adecuación climática en valores de presencia/ausencia en modelos de nicho climático.

En el *Capítulo 3*, se describe un nuevo modelo macroecológico de simulación computacional ("autómata") que utiliza un paisaje de celdas bidimensional proyectado sobre capas paleoclimáticas para explorar el papel jugado por la conservación del nicho de las especies, la aparición de barreras temporales a la dispersión, y la extinción asociada a cambios climáticos, en la formación de patrones de distribución espaciales disyuntos, como la Rand Flora. A diferencia de modelos anteriores, que consideran la formación de una especie por una tasa de especiación como un evento único e independiente del resto, el modelo propuesto permite incorporar la historia evolutiva de un linaje

mediante la generación de filogenias como resultado del modelo. La especiación se asume por una tasa de especiación, mientras que la probabilidad de extinción se hace dependiente de factores abióticos como el clima y bióticos, como el número de especies existentes en esa celda. La evaluación del ajuste del modelo se hace por comparación entre los patrones espaciales (rango geográfico) y evolutivos (número de especies por celda y estructura de la filogenia) generados por la simulación y los observados en estudios empíricos, como los del patrón biogeográfico Rand Flora, como por ejemplo en el género *Camptoloma*.

Nos enfrentamos a una gran crisis de la biodiversidad y a un calentamiento global en parte inducido por la actividad humana. Ante la imposibilidad de conservar toda la biodiversidad existente, se hace necesario desarrollar nuevos enfoques que incrementen la precisión de nuestras predicciones de la respuesta evolutiva de las especies. Esta tesis pretende dar cabida a este objetivo.

GENERAL INTRODUCTION

CLIMATE CHANGE AND BIODIVERSITY LOSS DURING THE CENOZOIC ERA

A changing climate

At virtually every temporal and spatial scale, drastic environmental changes have occurred during the Phanerozoic (Hannisdal and Peters 2011). The most widely documented environmental changes concern temperature and the rise and fall of sea levels (Zachos et al. 2001, 2008; Benton 2009; Barnosky et al. 2011a; Condamine et al. 2013). Four successive phases of warming and cooling events mostly characterise the Phanerozoic, with the most recent ending just after the Chicxulub "the dinosaur killer" meteor impact that initiated the K/T mass extinction event (Zachos et al. 2001; Purvis 2008; Condamine et al. 2013) and began the Cenozoic Era. These warming and cooling events are often linked to periods of intense tectonic activity that remodelled the Earth's configuration, changed major oceanic currents, and caused volcanic eruptions that released carbon dioxide into the atmosphere (Figure 1a, d, e, f; Zachos et al. 2001, 2008). The Cenozoic Era covers the last 65 million years of the Earth's history, and provides an ideal foundation from which to understand relationships betweean carbon cycling and climate (Zachos et al. 2008). Environmental changes during the Cenozoic are well documented (Figure 1; Miller et al. 2005; Zachos et al. 2001, 2008). In contrast to the present day, much of the early Cenozoic was characterised by noticeably higher concentrations of greenhouse gases, as well as a much warmer mean global temperature and poles with little or no ice (Figure 1a, b; Zachos et al. 2001; Royer 2006). During the entire Cenozoic era, absolute temperature had, on average, steadily decreased, with an increase in the final million years before present day (Figure 1a; Hansen et al. 2006).

The Paleocene and Eocene are separated by the Late Paleocene Thermal Maximum and marked by two main geological events: the rifting of the North Atlantic, and volcanism of the India-Asia contact (Figure 1c, e; Zachos et al. 2001). During the Early Eocene, 55 million years ago, the Earth experienced one of the warmest intervals in history, known as the Early Eocene Climatic Optimum, with global temperatures up to 10°c warmer than present and tropical climates characterising northern latitudes (Figure 1a, c; Zachos et al. 2008). The boreotropical flora, a mixture of hardwood deciduous and evergreen tropical taxa with no analogue today, extended across North America and Eurasia (Figure 1h, Tiffney 1985a, b; Wolfe 1975). Several early Eocene

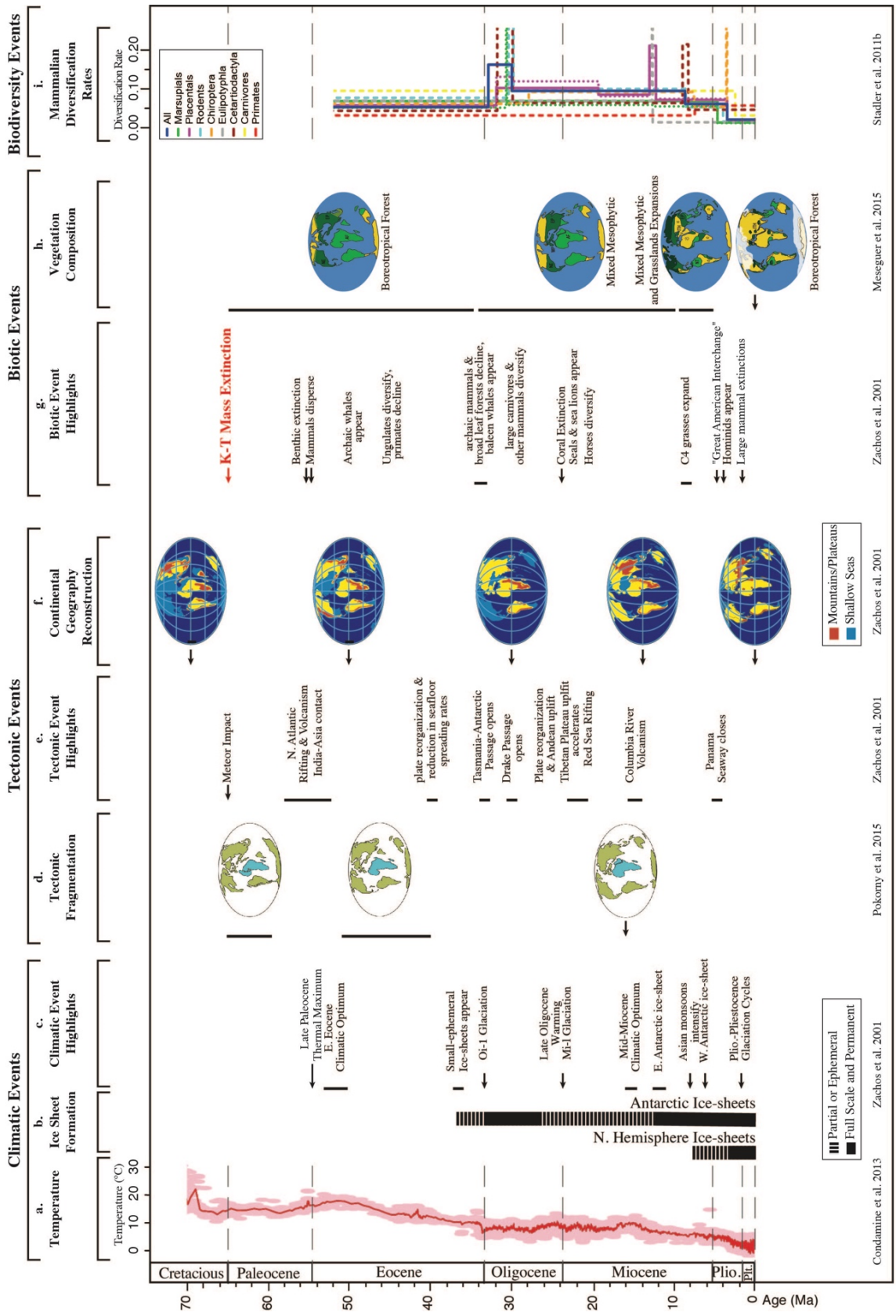


Figure 1. Diagram showing major climate change, geological tectonism, and biotic-biodiversity events that occurred during the Cenozoic, as recorded in the literature. a. Temperature: Relative proportions of different oxygen isotope ($\Delta^{18}O$) in samples of benthic **Figure 1** cont. foraminifer shells estimate major trends in global climate change during the Cenozoic (65 Ma to present day, Zachos et al. 2008). Absolute

Figure 1. cont. temperature was calculated from $\Delta^{18}O$ using Epstein et al. 1953's formula $T = 16.4 - 4.3 \times \Delta^{18}O + 0.14 \times (\Delta^{18}O)^2$. b. Ice Sheet Formation: The vertical bars provide a rough qualitative representation of ice volume in each hemisphere relative to the LGM, with the dashed bar representing periods of minimal ice coverage (50%), and the full bar representing close to maximum ice coverage (50% of present). d. Tectonic Fragmentation: Tectonic fragmentation of the supercontinent Gondwana through geological time. f. Continental Geographical Reconstruction: Continental geography reconstructed for five intervals of the last 70 Ma; the formation of shallow seas and mountains/plateaus is shown. h. Vegetation Composition: Scheme representing changes in vegetation composition in response to major shifts in Cenozoic climate based on paleobotanical evidence. The close-canopy tropical forest in the Southern Hemisphere is represented by light green colour over a yellow background. The northern vegetation belt in dark green colour represents, alternatively, the boreotropical forest (65-35 Ma), and its successors: the mixed-mesophytic forest (35-10 Ma, 10-3.5 Ma) and the temperate forest (18,000 years ago). The boreotropical belt in the northern regions of Eurasia and North America (10-3.5 Ma, 18,000 years ago) is represented by a darker shade of green; starbursts represent the expansion of grassland biomes after the Miocene. Abbreviations: NE (Nearctic), AF (Sub-Saharian Africa), EP (eastern Palearctic), WP (western Palearctic), ITH (Irano-Turanian-Himalayan), NT (Neotropic), OC (Oceania). i. Mammalian Diversification Rates: Maximum-likelihood diversification rate estimates (per million years) for the mammalian phylogeny and the mammalian subclasses placental (eutherian), marsupial (metatheria), and the six largest placental orders. The number of shifts is determined with the likelihood-ratio test at the 99% level. The dotted line for the marsupials corresponds to a 95% level; the dotted line for the placentals corresponds to a model allowing for one more shift than at the 99% level (which has a P value of 0.83). Some key climatic (c.), tectonic (e.) and biotic (g.) events are listed, as well. Information was obtained from the following sources: (Lawver and Gahagan 1998; Copeland 1997; Alroy at al. 2000); Stadler 2011b; Condamine et al. 2013; Meseguer et al. 2015.

hyperthermals, including the Eocene Thermal Maximum, have been documented (Figure 1a, c; Lourens et al. 2005). The Antarctic ice began to form as small-ephemeral ice sheets in the Late Eocene, just before the Oligocene 1 Glaciation event. The established ice sheets formed by this glaciation event lasting until right before the Late Oligocene Warming (Figure 1a, b, c; Zachos et al. 2001, 2008).

During the period of the Early Miocene, changes in global atmospheric circulation profoundly transformed the climate of the African continent (Griffin 2002), which would have limited biotic connections between northern and southern Africa (Sanmartín et al. 2010). Desertification of the African continent started in the southwest (17-16 Ma Mid Miocene), with the formation of the current Namib Desert, which subsequently advanced northeastwards, along with the expansion of savannahs in eastern Africa (8-9 Ma Late Miocene) and the appearance of the large Sahara Desert in North Africa (7 Ma Late Miocene; Senut et al. 2009 The early Miocene marked the climax of the continual and gradual northeastward drift of Africa from the supercontinent Gondwana, and its final collision with Eurasia (Figure 1d), signalling the closure of the Tethys Seaway (Terminal Tethyan Event, TTE; Liu et al. 2018) by the Arabian Plate in the mid-Miocene (18–12 Ma; Figure 1e). The TTE, preceded by the northward movement of the African–Arabian

plates and simultaneous counter clockwise rotation of the Arabian plate, formed the so-called Gomphotherium land bridge, which permitted terrestrial interchange between Eurasia and Africa during the early Miocene (Harzhauser et al. 2007; Kapli et al. 2015). It also separated the Atlantic Ocean from the Indian Ocean, leading to the vicariance or split of their marine faunas (Adams and Whybrow 1983). The late Miocene has attracted recent interest as a potential model system for testing future climate change scenarios (Lunt et al. 2008). Climate proxy data for the Tortonian (11.6–7 Ma) was characterised globally by: colder and wetter conditions (e.g. Asian monsoons, Figure 1c; Beerling et al. 2009), which were warmer and more humid conditions than those are today in continental Europe (e.g. Mosbrugger et al. 2005; Bruch et al. 2006); the expanses of the C4 grasses (Figure 1g); an already well-developed Antarctic ice cap (Figure 1b); and a mostly ice-free Greenland (Figure 1b; Thiede et al. 1998). The land-sea distribution was similar to present (Figure 1d), but probably with less pronounced topography and still open oceanic gateways that now are either completely closed (Isthmus of Panama Seaway closes, Figure 1e, f; deMenocal 2004; Böhme et al. 2008; Liu et al. 2018) or at least restricted for large-scale oceanic interexchange (Indonesian seaway; deMenocal 2004; Böhme et al. 2008; Liu et al. 2018). Overall, warmer ocean temperatures have been deduced from deep-sea proxy records (Zachos et al. 2001; Lear et al. 2003). Hence, it is expected that higher sea-surface temperatures in the North Atlantic, and an enhanced northward heat transport and moisture supply from low to high latitudes, led to an intensified hydrological cycle in Europe (Böhme et al. 2008).

The Early Pliocene was a relatively humid and colder period. The East Antarctic ice-sheets had already re-established in the late Miocene (Figure 1b). However, during this period the West Antarctic ice-sheets were also re-established, along with the initial formation of the Northern Hemisphere ice-sheets (Figure 1b; Zachos et al. 2001, 2008). The Mid Pliocene Warming Event was a consequence of global (e.g. the closing of the Panama Isthmus, Figure 1a, e; Zachos et al. 2001) and regional tectonic events (e.g. the uplift of the East Africa Rift System, Figure 1f), which created a rain shadow between central-west Africa and the drier east African plateau (Sepulchre et al. 2006). The Plio-Pleistocene Glaciation Cycles began upon entering the Pleistocene era (2.6 Ma, Figure 1c).

Is biodiversity loss accelerating?

Some studies (such as Pimm et al. 2006, Pereira et al. 2010 and Barnosky et al. 2011a) show that present day background extinction rates are higher than what would be expected from the fossil record inferences, indicating that the Earth's sixth mass extinction event may have already arrived.

Analyses of cetacean (whales and dolphins) background extinction and speciation rates illustrate that within-clade diversity has been declining for the last 5.3 million years, and furthermore that this decline is nested within an even longer-term decline that began some 14 million years ago (Barnosky et al. 2011a; Morlon et al. 2011; Condamine et al. 2013). Diversification rates have also been estimated across two of the three mammalian subclasses, with these rates shown to be declining. Stadler et al. 2011b estimated diversification rate shifts for the: placentals (eutherian) and marsupials (metatheria), as well as for the six largest placental orders: rodents, chiroptera (bats), eulipotyphla (shrews, moles, hedgehogs), cetartiodactyla (whales), carnivores, and primates (Figure 1i; Stadler 2011b). Stadler (2011b) inferred a significant increase in diversification rates for rodents and cetartiodactyla in the placentals at the beginning of the Oligocene; however, only rodents and cetartiodactyla (placentals) and marsupials showed a peak. Moreover, Stadler (2011b) demonstrated that all analysed groups of mammals, excluding eulipotyphla, cetartiodactyla and primates, have a significant decrease in diversification rates during the last million years.

It has been estimated that current background extinction rates are higher than those that caused the Big Five extinctions (Figure 2) in geological time (Pimm and Joppa 2015; De Vos et al. 2015), and that if these rates were to continue as they are now, they could create extinction magnitudes as large as those of the Big Five in as little as three hundred years (Barnosky et al. 2011a). Currently, there is suggestion (Huang et al. 2015) that there is a disproportionate loss of evolutionary history happening now in the Anthropocene (a proposed new geological era that succeeds the Holocene, defined as the part of the Quaternary period that which human activity has been the dominant influence on climate and the environment). Although, as to how much of the evolutionary history loss is human-led, anthropogenic-induced extinction, and how much of it is just natural cycling of diversification rates is unclear (Moen and Morlon 2014; Huang et al. 2015). There is a great need for future research, for much is still unknown in terms of understanding: whether "critically endangered", "endangered" and "vulnerable" species will for definite become extinct; whether the present day background extinction rates calculated will continue their upward trajectory, or whether they will decrease; and how reliable the background extinction rates in well-studied taxa (e.g. Morlon et al. 2011; Stadler et al. 2011b; Condamine et al. 2013) can be extrapolated to other species (Baillie et al. 2008; Mace et al. 2008; Hoffmann et al. 2010; Pereira et al. 2010; Barnosky et al. 2011a). So far, exploration as to whether present-day background extinction rates within lineages fall outside expectations with respect to long-term diversity dynamics has been limited. This makes it difficult to know whether a "threatened" genus is indeed truly in threat of becoming extinct, or whether it is just naturally cycling through a period of low diversification. Knowing these long-term

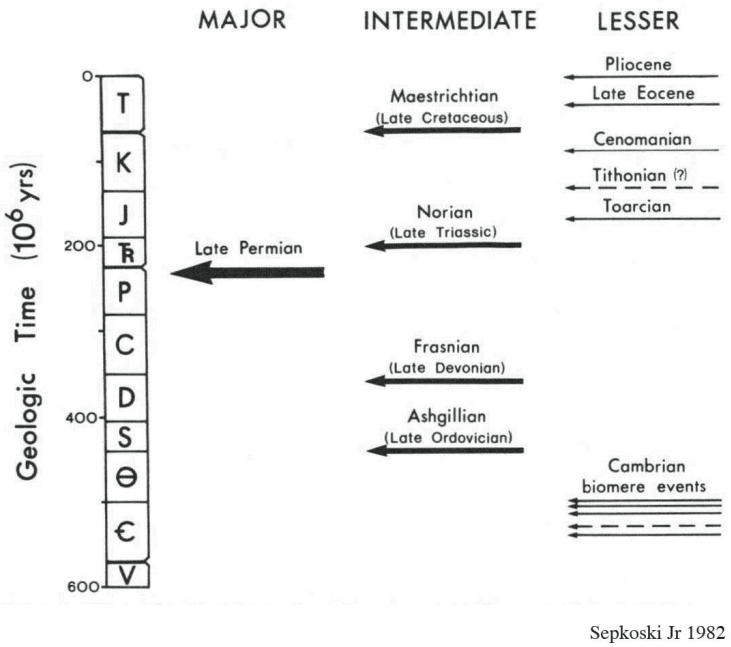


Figure 2. Scheme showing major, intermediate, and lesser mass extinction events through geological time as defined in Sepkoski, 1982. Dashed arrows and "?" indicate poor documentation of the mass extinction event. The geologic systems are indicated and denoted by standard symbols with "V" standing for Vendian. Taken from Sepkoski (1982).

diversification dynamics would allow us to know whether the genera is experiencing a background extinction rate that is an order of magnitude higher than anything it has experienced during its evolutionary history. This knowledge would therefore allow for a more accurate classification of "threatened", even if this global extinction could take up to 100,000 years to occur (Barnosky et al. 2011a; Quental and Marshall 2010; Huang et al. 2015), and hence aid in the decision as to where conservation priorities should focus –should they focus, for example, on species-poor, geologically old lineages as suggested by Huang et al. 2015, or on the great green wall project that aims to create the largest living structure on Earth that will help combat African aridification and global climate change (www.greatgreenwall.org)?

INFERRING SPECIES RESPONSES UNDER THE INFLUENCE OF CLIMATE CHANGE

The current concern on anthropogenic-induced climate change and its impact on biodiversity levels, through land degradation and dramatic changes in the environmental landscape, has increased the interest in reconstructing past species' responses to climatic events. These include: adaptation

(evolving a new trait) that eventually can result in speciation; persistence in their geographical locations (via genetic and phenotypic plasticity); geographic range shifts (displacement, expansion and contraction); and extinctions (local, global and mass extinction events) (Diniz-Filho and Binin 2008; Willis and McDonald 2011; Araújo et al. 2013; Romdal et al. 2013; Meseguer et al. 2018). In Africa, phylogenetic and fossil data suggest that plants responded in different ways to the aridification trend that started with the closing of the Tethys Seaway in the Early Miocene, from extinction and geographic range shifts in tropical and temperate-adapted lineages (Plana 2004; Pokorny et al. 2015), to rapid diversification driven by allopatry and adaptation in other groups (Thiv et al. 2010; García-Aloy et al. 2017; Pirie et al. 2018). In this thesis, I have focused mainly on two of these organismal responses to climate change: extinction and geographical range shifts; the two are closely interconnected.

Evolutionary episodes of hyperdiversification –i.e. speciation rates within a lineage that are significantly higher than those expected under the background diversification rate of their encompassing clade– have long attracted the attention of biologists (Hughes and Eastwood 2006; Valente et al. 2010) and analytical systematists (Rabosky 2006, 2014; Alfaro et al. 2009) because of their potential links to "key innovations" (morphological novelties) or the colonisation of novel environments leading to increased species fitness, "key opportunities" (Wiens et al. 2010; Donoghue and Sanderson 2015). By contrast, episodes of high extinction rates such as mass extinction events (MEE), have traditionally received less attention in the phylogenetic literature because of the difficulty in measuring a process that removes rather than generates diversity (Pyron and Burbrink 2012; Sanmartín and Meseguer 2016). As many as 15 separate MEEs have been identified in the Phanerozoic marine fossil record (Figure 2; Sepkoski 1982), and are classifiable under three different categories: major, intermediate and lesser, with the major events affecting almost all lineages across the world and the intermediate and lesser affecting only a few localised lineages (Sepkoski 1982). MEE are distinguished in the paleontological records as widespread, higher taxonomic group extinctions (e.g. up to 96% of marine invertebrate species became extinct during the largest, late Triassic MEE; Raup 1979). A MEE is defined as a period: (i) where the ratio of the background extinction rate over the speciation rate, i.e. the turnover rate is larger than 1; and (ii) during between less than one million up to about 15 million years (Mya), a time interval dependent on the magnitude or intensity of the MEE (Sepkoski 1982). Unlike speciation or background extinction rates, which are assumed to depend on species biotic traits or a clade's ecology (Purvis 2008; Ezard et al. 2011), MEEs are monocyclic (irregularly cycling) and are often linked to abiotic factors, i.e. long-term environmental changes or catastrophic, geological events whose effects are felt across many lineages

(Sepkoski 1982; Jablonski 2008; Pyron and Burbrink 2012; Sanmartín and Meseguer 2016), and hence form a key element of the paleontological record, responsible for major ecosystem reordering and change (Raup 1979; Sepkoski 1982; Gould 1994; Purvis 2008; Benton 2009). MEEs are often compared to current-day greenhouse-induced climate change (IPCC 2001) and recently have regained importance in the context of human-induced biodiversity loss (Barnosky et al. 2011a, b). In Chapter 1, I extend a method, the Bayesian Skyline Birth-Death model, which has been developed within an epidemiology context, but is here applied to the inference of rapid increases of extinction rates (mass extinction events) over geological time from extant diversity data.

Geographical range shifts are often predicted as the outcome of rapid climate change, under the assumption that the rate of environmental change is greater than that of species adaption (Martínez-Meyer and Peterson 2006; Thuiller et al. 2006; Waldron 2010). During climatically adverse periods, species may survive by contracting their geographical range as they track their niche ("climatic refugia"), from which they disperse, once conditions become more favourable, hence re-expanding their geographical range. The tracking by a species of its niche and the failure to adapt to environmental oscillations effects within its geographical range have been characterised as niche conservatism (Wiens 2004). When a change in the physical template occurs that is too large for populations to migrate or adapt, fragmentation of the species' original distribution into smaller, spatially disjunct ranges and subsequent allopatric speciation (Wiens 2004) or even extinction (McDonald and Brown 1992) may result. Reproductive isolation and reduced gene flow among these disconnected ranges may lead to larger inter-population genetic differences, and finally, to speciation (Dorn et al. 2014). However, range contractions can also result in smaller population sizes, reduced genetic diversity and a higher extinction risk, the "extinction debt" (Waldron 2010; Mairal et al. 2018). In Chapters 2 and 3, I explore new avenues for uncovering the evolutionary signature of climate change on species and organism geographical ranges over time, using two different approaches: i. biogeographic analysis combined with ecological niche modelling, and ii. a novel forward-simulation spatially explicit ("in silico") method.

Below, I describe in more detail the background and novel insights provided by these approaches.

INFERRING MASS EXTINCTION EVENTS FROM EXTANT ONLY PHYLOGENIES

Traditionally, MEEs have been studied from paleontological evidence (Raup 1979; Sepkoski 1982; Jablonski 2008). If a clade is associated with a sufficient fossil record, then this record can be

used to quantify the most probable number of MEEs within the specific time period that spans the fossil data (Raup 1979). However, most clades either possess an incomplete fossil record or lack a fossil record entirely. High extinction rates, as those associated with MEEs, can also leave an imprint on the timing and structuring of cladogenetic events in phylogenetic trees containing only extant taxa (Harvey et al. 1994). Macroevolutionary birth-death models can be used to find out whether a species' past response to environmental change was to adapt or become globally extinct, solely through the estimation of changing speciation and background extinction rates on the species' phylogenetic tree. These models have been adapted to be used to detect and locate MEEs in phylogenies that possess a complete, an incomplete, or no fossil record. They are all based on the birth-death framework described below.

i. The birth-death framework

Regarding macroevolution, a birth-death model is one that is used to describe how the number of species within a phylogeny changes through time, with the model returning an estimation of the rate of a "birth" or "speciation" event (λ) and a "death" or "background extinction" event (μ) occurring within the phylogeny. Here, each species has an equal probability of speciating or becoming extinct and only one species may speciate or become extinct at any one time (Nee et al. 1994). The simplest null model to estimate these rates is the constant birth-death model, which assumes that λ and μ are constant over time and complete sampling of extant diversity at present time. The Yule model is a special case of the constant birth-death model, where μ equals zero. Over the last decades, increasingly more sophisticated and realistic birth-death models have been developed, from the initial constant birth-death model to cope with violations of the constant-rate birth-death model: i. to deal with reconstructed phylogenies that contain limited or lack extinct taxa (introduced by Nee et al. 1994); ii. to accommodate, through a sampling parameter, incomplete extant species sampling: either a random fraction of missing species in the phylogeny, or clades collapsed to single tips in the phylogeny due to higher taxonomic sampling level; Höhna et al. 2011; Stadler and Bokma 2012); iii. to deal with time-variable, "discrete" birth-death models, in which speciation (λ) and background extinction (μ) rate estimations are assumed constant within a time interval but can, experience rate shifts between time intervals in the phylogeny, or punctual sampling events in the past, equivalent to a mass extinction event (Stadler et al. 2011b, described below); iv. to accommodate time-varying (continuous) variation in λ and μ (e.g. Morlon et al. 2011); and v. to test

whether rates of λ and μ are dependent on an environmental variable, itself varying over time, e.g., temperature or elevation (e.g. Rabosky and Lovette 2008; Condamine et al. 2013, 2018).

Antonelli and Sanmartín (2011) presented a framework that mechanised Harvey et al. (1994) and Crisp and Cook (2009)'s suggestion that phylogenetic trees, simulated under alternative diversification birth-death models, could help understand how macroevolutionary processes (such as punctual high extinction rates, i.e. MEEs, and high relative μ) leave their signature upon the the temporal pattern of individual lineage diversification when their lineage-through-time (LTT) plots are compared to those of empirical phylogeny LTT plots. With this framework, Antonelli and Sanmartín 2011 explored whether the large temporal gap found in the empirical Chloranthaceae (genus *Hedyosmum*) phylogeny was the signature of MEEs, scenarios of constant high μ, or a combination of the two, occuring at 65 Ma (K/T) and 36 Ma (Late Eocene MEE). Antonelli and Sanmartín (2011) concluded that while comparing the shape of LTT plots from reconstructed phylogenies alone was not enough to discriminate between a MEE and high relative μ scenario, especially when faced with incomplete taxon sampling, this framework provided an interesting concept of how to use birth-death models as a way of identifying and locating MEEs within a phylogeny. Despite claims to the contrary (Rabosky 2010), it is mathematically possible to estimate the background extinction rate from a reconstructed phylogeny with no fossil or extinct lineages, using the expected departure in the distribution of origin times (LTT plot) under a constant-rate birth-death model with positive extinction against a pure-birth model (i.e., the "pull of the present"; Nee et al. 1994; Sanmartín and Meseguer 2016). The problem arises because diversification rates rarely follow a constant-rate birth-death model. Variation in diversification rates over time and/or across lineages leads to underestimation of extinction rates from extant only phylogenies. For example, if one clade is declining over time, with μ higher than λ, but it is otherwise included within a larger clade, with other clades expanding in diversity over time, the overall diversification rates is often (erroneously) estimated as constant-rate (Quental and Marshall 2010). Same if there are frequent diversification rate shifts over time, which are smoothed over as a constant-rate pattern (Sanmartín and Meseguer 2016).

Inspired by Antonelli and Sanmartín (2011)'s work, Stadler (2011b) introduced a birth-death model that was able to detect the presence of rate shifts based on phylogenetic trees containing extant taxa only. This model assumes discrete time intervals during which λ and μ are constant, and the rates may change arbitrarily between intervals (Stadler 2011b). Unlike previous attempts at using a likelihood framework to identify rate shifts in diversification, by partitioning the phylogeny into "before" and "after" time intervals (Rabosky, 2006), Stadler's whole-tree birth-death models could

discriminate between the pull of the present and past rate shifts. It was implemented within a maximum likelihood (ML) framework in the R package *TreePar* (Stadler 2011b), and used successfully to detect the timing of MEEs in phylogenies that have a large to moderate number of terminals (e.g. N > 200-500 taxa; Laurent et al. 2015; Beaulieu and O'Meara 2015). Sanmartín and Meseguer (2016) found that this model underperforms with relatively small phylogenies (N < 50 taxa).

Stadler's whole-tree birth-death model can also be used to detect the phylogenetic signature of mass extinction events (MEEs). Unlike speciation or background extinction rates, which are assumed to depend on species biotic traits or a clade's ecology (Purvis 2008; Ezard et al. 2011), MEEs are often linked to abiotic factors (Pyron and Burbrink 2012), and so, MEEs are often modelled in the phylogenetic literature as tree-wide events that act simultaneously across clades in contrast to events that are clade-specific (Sanmartín and Meseguer 2016). Stadler (2011a, b) model MEEs as a random instantaneous extinction event, in which a significant fraction of the standing diversity is instantaneously and simultaneously removed from the phylogenetic tree at a specific point in time t. The fraction (number or percentage of species) that gets extinct is controlled by the magnitude of the MEE, with the magnitude or intensity of the mass extinction, being defined as 1 minus the survival probability of each species at the MEE (ρ) = $1 - \rho$ (Harvey et al. 1994; Stadler 2011b; May et al. 2016). This model to describe MEEs as instantaneous events of sampling intensity is called the “single pulse” MEE model, and is the one used more often in mathematical birth-death modelling (Harvey et al. 1994; Stadler 2011b, c; Laurent et al. 2015; May et al. 2016).

For phylogenetic trees spanning millions of years, it is likely they have been affected by rate shifts and single-pulse MEEs, perhaps caused by global (climatic or geological) events (Laurent et al. 2015). However, though methods such as Stadler (2011b) can be used in principle to estimate the timing and magnitude of single-pulse MEEs from extant only taxa, it remains difficult to simultaneously estimate the frequency of tree-wide rate shifts in diversification and single-pulse MEEs due to issues of parameter non-identifiability, i.e. when different combinations of parameter values yield flat likelihood surfaces for part of the parameter space (Rannala 2002). In fact, under a ML framework, it remains impossible to distinguish between a constant birth-death process with single-pulse MEEs and a process in which diversification rates vary discretely over time because both types of processes generate identical phylogenetic signatures and have comparable likelihood functions (Stadler 2011b; Sanmartín and Meseguer 2016). This is known as overparameterisation. Hence, in Stadler (2011b)’s method *TreePar*, one of these parameters must be fixed, for example, by assuming that μ and λ have remained constant before and after the single-pulse MEE event, or by

fixing the intensity of the MEE before inferring the timing and number of rate shifts (Stadler 2011b). Also, the algorithm cannot estimate multiple rate shifts simultaneously; instead, it uses a greedy approach, where the time of one rate shift is estimated and fixed before estimating the time of the next rate shift (Stadler 2011b).

To solve this problem, May et al. (2016) introduced a Bayesian statistical inference approach to the single-pulse MEE model, the Compound Poisson Process (CPP) on Mass Extinction Times (CoMET), implemented in the R package *TESS* (Höhna et al. 2015). Bayesian inference is less problematic under overparameterisation than equivalent likelihood-based approaches due to the integration of parameter uncertainty through estimation of marginal likelihoods. CoMET implements a stochastic branching process model in which rates of speciation and background extinction are constant between rate shifts, and single-pulse MEEs are modelled as tree-wide instantaneous extinction events. Specifically, the method considers three types of events: instantaneous tree-wide shifts in speciation rate, instantaneous tree-wide shifts in background extinction rates, and instantaneous tree-wide single-pulse MEEs. Each of them is modelled through a separate CPP, with waiting times distributed exponentially according to event-specific rate parameters (May et al. 2016). To address the problem of parameter non-identifiability in single-pulse models, CoMET implements a hierarchical Bayesian approach in which rate shifts in λ and μ are considered as "nuisance" parameters that are integrated over in the estimation of the marginal posterior probabilities of the focal parameters: the time, number and the intensity (magnitude) of single-pulse MEEs. However, CPP models themselves are very sensitive to the choice of priors, which means that in practice some parameters of the model, such as the magnitude of the single-pulse MEE, are often assigned informative empirical priors (May et al. 2016).

Another serious problem with the single-pulse MEE model is the fact that it considers MEE as instantaneous events, with intensity but without time duration (Figure 3a). This definition of MEEs stands in contrast with the one used by palaeontologists, which characterise mass extinctions as times when the Earth loses more than three quarters its species (75% to 99% intensity) in a geologically short interval which oscillates between 1 and 15 million years (Sepowsky 1982; Benton 2009; Barnosky et al. 2011a, b). Condamine et al. (2013) proposed instead a "time-slice" model, in which MEEs are modelled as times when a significant increase in μ for a short time interval, with turnover rate $\varepsilon = \frac{\mu}{\lambda} > 1$, is followed by a decrease in μ that returns to $\varepsilon < 1$ in the next time interval (Figure 3b). In Chapter 1 of this thesis, I propose a new approach in mathematical birth-death modelling that allows to infer MEEs as period of times when the background extinction rate spikes higher than the speciation rate, and above the background level, with at least 75% depletion of the

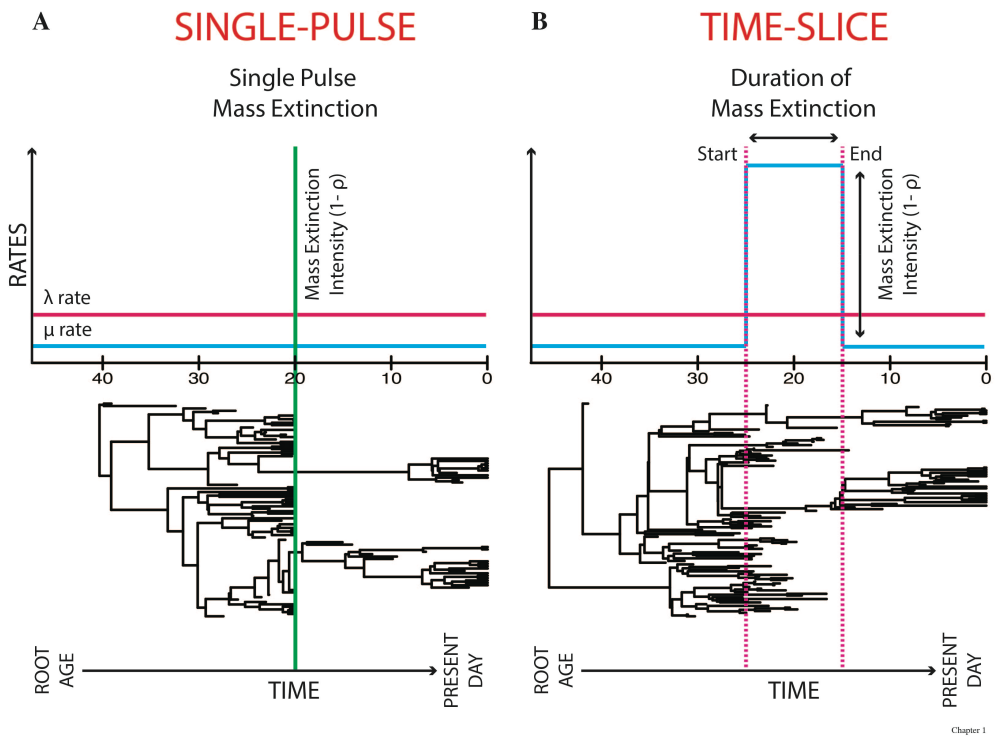


Figure 3. Two different mass extinction event (MEE) models used in the birth-death framework. Left: The "single-pulse" scenario models MEEs as a past sampling event in which a significant percentage of species is simultaneously and instantaneously removed from the phylogenetic tree, at a specified time. Right: The "time-slice" scenario models MEEs as a significant increase in the extinction rate, μ, for a specific period of time, where the turnover or background extinction rate, $\varepsilon = \frac{\mu}{\lambda} > 1$, followed with a decrease in μ that results in a return to ε < 1; λ is assumed to be unchanged before and after the MEE. In the two trees, the "pre-MEE" μ is equal to "post-MEE" μ but this is not necessary. The "single-pulse" scenario is the model used by most birth-death methods (Stadler 2011b; May et al. 2016); the "time-slice" scenario is the one explored in Chapter 1, through the use of a Birth-Death Skyline model.

extant diversity (Barnosky et al. 2011a, b). Specifically, I borrow from the epidemiological literature the Birth-Death Skyline model, which was developed to detect bottlenecks in infection dynamics (Stadler et al. 2013), but is here tested for detecting the phylogenetic signature of MEEs under a time-slice approach.

Inferring Ancestral Plant Distributions

Traditionally, inferences on ancestral palaeoplant distribution were built on evidence including: i. secondary inferences of faunal assemblage (e.g. Pickford and Senut 1999, 2003; Ségalen et al. 2006, 2008; Senut et al. 1997, 2009); ii. the fossil record, such as pollen and seed fossils (e.g. Dechampes 1984; Dechampes et al. 1992; deMenocal 2004; De Wit and Bamford 1993) and

fossilised fauna teeth (e.g. Cerling et al. 1997; deMenocal 2004); iii. stable carbon isotopic composition of paleosol carbonates, used as an indicator of C4 photosynthesis in modern and ancient ecosystems, and therefore the composition of C4 plants present (e.g. Cerling et al. 1989; Cerling 1984, 1992; Quade et al. 1989a, b); iv. stable carbon isotopic composition to infer C3 plants (e.g. Deines 1980; Ehleringer 1988); and v. organic geochemical analytical techniques for molecular biomarker analyses of plant leaf waxes (n-alkanes) preserved in deep-sea sedimentary organic matter (e.g. Schefuss et al. 2003).

Palaeovegetation composition maps (e.g. Figure 1h), built from the descriptions of ancestral palaeoplant distributions, are neither greatly specific nor detailed because of the paucity of plant palaeontological studies (Senut et al. 2009). Only general terms such as “savannah”, “grasslands” and “dwarf shrubs” are used. The definitions of vegetation composition vary greatly, for example, the term “savannah”, although useful for describing open tropical vegetation, encompasses a wide range of vegetation types and has different meanings on different continents (Cole 1986; Pratt et al. 1966; Sarmiento 1984). The plant species that may be included in these maps are limited due to differences in the quality of the fossil record among species and lineages. Lastly, the geographical space that these maps cover is limited to those areas where some fossil evidence has been found. On the other hand, obtaining accurate and detailed ancestral palaeoplant distributions and palaeovegetation composition maps is crucial, not only as primary evidence of relationships between plant species and climatic change with respect to their evolutionary and distributional history, but also as secondary evidence in understanding possible relationships between faunas and climate change (e.g., deMenocal 2004; Leakey et al. 1996; Pickford 2008; Senut et al. 2009).

In recent decades, biogeographic inference models and ecological niche model (ENM) approaches have taken a prominent role in bringing a new mechanistic understanding of ancestral palaeoplant distributions. Rather than relying on the fossil record, these approaches draw upon other sources of evidence, such as DNA sequence data and presence/absence location information of extant, and when possible, extinct species. Hence, their power to infer ancestral palaeoplant distributions is not hindered by, for example, lack of fossil records or C3/C4 analyses, as more traditional methods can be. These methods are also able to provide more detail into distribution inferences, which can be lineage specific unlike paleostratigraphic approaches. Furthermore, both biogeographic inference and ENMs can also be used to understand the evolutionary signature of changes within species niches for modelling species range dynamics (Svenning et al. 2015).

Biogeographic inference: Correlating distribution patterns to phylogenetic patterns

In the genomic era, as large-scale DNA sequencing has become increasingly cheaper, and computers powerful enough to analyse thousands or hundred thousands of DNA sites and generate robust phylogenetic trees, the use of a biogeographic inference framework has become a popular and viable approach to reconstruct ancestral plant distributions. In biogeographic inference, extant, and when possible extinct, geographical distribution information, together with a phylogenetic hypothesis specifying evolutionary relationships among species, are supplied, and a statistical model is then used to reconstruct the evolution of geographical ranges back through time upon the ancestral nodes of the phylogenetic tree (Ree and Smith 2008; Ronquist and Sanmartín 2011; Mairal et al. 2015). The result is an understanding of when and where adaption or geographical range shift events took place. If this tree is time calibrated, it is also possible to date these events. Biogeographic inference models, however, typically do not consider detailed spatially explicit information: geographical distributions are defined as abstract "arbitrary" areas, without geographical coordinates. These models work on the belief that taxa that share similar phylogenetic and distributional patterns must also share a common biogeographic history (Ronquist and Sanmartín 2011) and hence, lack a geographically meaningful framework (but see Tagliacollo et al. 2015). Also, biogeographical inference lacks an ecological framework, meaning that it is not required to know past species' climatic ecological preferences to infer ancestral ranges or the sequence of geographical events (but see Sanmartín et al. 2001 for an exception). This combination of the lack of a geographical and ecological framework means that there are no assumptions on the conservation or evolution of the species fundamental niche within the phylogeny, which makes for a simpler model.

In Chapter 2 of this thesis, I used three different likelihood-based biogeographic inference approaches, all of which employ discrete-state continuous time Markov Chain processes to model geographic range evolution (Ronquist and Sanmartín 2011) and are implemented onto the Bayesian framework: i. Dispersal-Extinction-Cladogenesis (Ree and Smith 2008) model, implemented in the Bayesian software RevBayes (Höhna et al. 2016), as described in Landis et al. (2018); ii. Bayesian Discrete Trait Analysis approach (Lemey et al. 2009) implemented in the software BEAST 1.8; and iii. Bayesian Structured Coalescent Approximation (De Maio et al. (2015), implemented in BEAST v2.4.7 (Bouckaert et al. 2014).

The Dispersal-Extinction-Cladogenesis (DEC; Ree and Smith 2008) model is a stochastic model that specifies instantaneous transition rates between discrete states (ranges) along phylogenetic branches, and applies them to estimate likelihoods of ancestral states (range inheritance

scenarios) at cladogenesis events (Ree and Smith 2008). DEC is a refinement on Ree et al. (2005)'s model and was presented to bridge a conceptual gap in phylogenetic biology between models and inference methods for character evolution and those for geographic range evolution (Ree and Smith 2008). Species geographic ranges are coded as a combination of presences/absences of the species in predefined discrete areas. While defining geographic ranges in such a flexible way is advantageous, it is also computationally intensive, since the possible number of ranges increases exponentially if all of the presence/absence combinations are allowed. This means that the analysis performed by the DEC model becomes intractable for more than ten areas (Ree and Smith 2008; Landis et al. 2013).

The Bayesian Discrete Trait Analysis approach (DTA, Lemey et al. 2009) –akin to the Bayesian Island Biogeography (BIB) model developed by Sanmartín et al. (2008)– uses Bayesian inference and a nucleotide evolutionary model to infer the history of migration events between individuals and single DNA copies. Markov Chain Monte Carlo (MCMC) simulations are used to simultaneously estimate the posterior distribution of phylogenetic relationships, branch lengths and geographic ancestral states, while accounting for uncertainty in all of these parameters, including the estimation of ancestral frequencies for the root (Lemey et al. 2009). Bayesian Stochastic Variable Selection (BSVS) can be applied to this model to find the dispersal pathways with non-negligible support (Lemey et al. 2009). Geographic ranges of species are assumed to include only one of the discrete states defined in the model at any one time, with the discrete states being the areas or geographic locations of the sequences. Transition rates between states and migration rates between areas are also reconstructed.

The Bayesian Structured Coalescent Approximation (BASTA, De Maio et al. (2015) is the answer to De Maio et al. (2015)'s criticism of DTA. De Maio et al. (2015) criticised DTA because of the unrealistic treatment of the migration-mutation process. Since the effect of migration on effective population sizes is not modelled in the likelihood of the coalescent (i.e. the dataset is assumed to belong to a constant-size, panmictic population), posterior probability values on ancestral inferences tend to be overestimated and the method is highly sensitive to unequal sampling effort among areas (De Maio et al. 2015). BASTA uses the structured coalescent process (MTT model, Vaughan et al. 2012) to model migration but is computationally more efficient in handling a large number of populations/areas.

The lack of a geographic framework causes three important limitations to biogeographic inferences. Firstly, biogeographic inferences are very much reliant upon the priori area information that is provided. If a priori area or distribution assigned to a tree node is false (e.g. there was a

difficulty in distinguishing between non native and invasive ranges), biased or incomplete (e.g. sampling effort differs between regions or countries), or the full distribution of an extant species is unknown, then the areas can be lost from the analysis entirely. Likewise, if the a priori assigned areas fail to cover all regions that formed part of the ancestral species distributions, the latter will be (unknowingly) reconstructed as incomplete, or even wrongly inferred or with error (Saupe et al. 2017). This is because in biogeographic inference areas can only be lost but not created new. The result of using incompletely defined a priori areas is to inflate estimates of evolutionary change (Saupe et al. 2017). Secondly, the level of detail that biogeographic inference can provide is reliant on the quality of the phylogenetic tree used for the reconstruction. Since biogeographic inference is limited to the nodes and tips of the tree, if the phylogenetic tree is sparse, with long stems and young crown groups ("broom-and-handle" shape), large periods of time within the tree will be unaccounted for, simply because these periods of time do not include any speciation or diversification events. Also, high background extinction rates (especially if area-dependent, Meseguer et al. 2015) can mislead biogeographic reconstructions through the selective removal of speciation events from the phylogenetic tree (Sanmartín and Meseguer 2016). Lastly, biogeographic inferences can only be used to obtain past projections of species distributions. It is not possible to use these models for forecasting upon the same landscape, or projecting upon a different landscape within the same or different time periods to investigate the effects of climate change in present-day distributions (Peterson et al. 2011).

Ecological niche models: Correlating distribution patterns to ecological patterns

As depositing compiled datasets into open source databases (e.g. WorldClim, worldclim.org; GBIF, https://www.gbif.org) is becoming easier and nearly mandatory for research publication, the use of Ecological Niche Models (ENM, see Owens et al. 2011 for distinction between ENM and species distribution models, SDMs) has also become popular. ENMs estimate the ecological preferences (fundamental niche) of a species based on its current and/or past (i.e. fossil evidence, Metcalf et al. 2014) distributional range and the environmental variables present in this range (e.g. temperature, precipitation, solar radiation and altitude; Kozak et al. 2008). Specifically, ENMs return the probability of a species being present at a specific locality, known as "habitat suitability values". These habitat suitability values can then be projected: i. spatially, to find the fundamental niche of a species when occurrence data is scarce or limited (Pearson et al. 2007; Hengl et al. 2009); or ii. temporally, over palaeoclimate scenarios (see Figure 5 as an example; Araújo and New 2007; Evans

et al. 2009; Meseguer et al. 2015; Mairal et al. 2017) or predicted future climate scenarios (e.g. Araújo and New 2007; Thuiller et al. 2008; Al-Quaddi et al. 2017; Carboni et al. 2018), to explore a species distribution in the past and the potential future, thus providing a better understanding how the species respond to climate change (Peterson et al. 2011).

ENMs can be characterised as "distances" or "curve fitting models". Distances, such as Ecological Niche Factor Analysis (ENFA, Engler et al. (2004) or Mahalanobis Distance (MD, Clark et al. 1998), can be used when only presence data is available (Clark et al. 1998; Basille et al. 2008; Hirzel et al. 2002; Meseguer et al. 2015), while curve fitting models, such as the generalised linear model (GLM) or the generalised additive model (GAM), require presence and absence data (e.g. Thuiller et al. 2006, 2008, 2009; Hengl et al. 2009; Mairal et al. 2015, 2017). For curve fitting models, it is quite often the case that real data for presence information is incomplete or biased towards certain regions, countries or species, due to sampling effort. To overcome this drawback, pseudoabsences are typically generated. The treatment of pseudoabsences is an important issue that can mislead ENM projections. Chefaoui and Lobo (2008) provide an insightful discussion on how the pseudoabsence selection method chosen greatly influences the ENM's percentage of explained variability, the scores of the accuracy measures, and, most importantly, the degree of constraint in the distribution estimated.

In Chapter 2, I used Hengl et al. (2009)'s ENM method, an extension of the approach proposed by Engler et al. (2004), which addresses the issues highlighted by Chefaoui and Lobo (2008) on the use of pseudoabsences in ENMs. This method combines the best characteristics of curve fitting models –point pattern analysis (kernel smoothing) and logistic regression kriging– with distance-based models (ENFA), into a single framework, to predict habitat suitability from occurrence-only data (code available at spatialanalyst.net /wiki/index.php/ Species_Distribution_Modelling; Hengl et al. 2009). First, point pattern analysis and ENFA were used to generate habitat suitability values from presence only data. These ENFA habitat suitability values were then used to weigh the selection of pseudo absences to create a "presence/pseudoabsence" data needed for generating the regression-kriging GLM models, which in turn was used to predict species habitat suitability across geographic space (Hengl et al. 2009).

Once the habitat suitability values have been generated, the common practice is to run these values through a truncation threshold. Truncation thresholds are used to transform the continuous habitat suitability value predictions of ENMs into presence/absence data in order to establish a species geographic range (Diniz-Filho et al. 2010). Typically, truncation thresholds are selected

using model fit tests, such as the Kappa statistic (Monserud and Leemans 1992; Fielding and Bell 1997), the Area Under the Curve (AUC) or the Receiver Operator Characteristic (ROC) curves. Recent methods such as the true skill statistic (TSS, Allouche et al. 2006) or the leave-one-out test (Pearson et al. 2007) were developed to compensate for potential shortcomings in the Kappa (Allouche et al. 2006) and the AUC methods (Lobo et al. 2008). All these approaches rely on using a subset of the original data, the "test or training dataset", to evaluate model fit through examination of true and false predictions of species presences and absences (Liu et al. 2005; Jiménez-Valverde and Lobo 2007). When the original data is complete and unbiased, these methods of model fit choice are reliable.

Typically, ENMs are easy to build because there is bountiful information on geographically meaningful coordinates of presence-only or presence and absence records as biological input, and of present-day global climate layers as environmental input. However, there are four important limitations faced by projecting ENMs temporally. Firstly, the hindcasting of ENMs upon palaeoclimate data and the forecasting upon future predicted climate data is currently very limited. This is because there are not many available (and reliable) climatic projections. For forecasting ENMs, there are four future climatic data projections from global climate models for greenhouse-gas scenarios for the year intervals 2041-2060 and 2070-2060 (accessible from www.worldclim.org/cmip5v1). For hindcasting ENMs, there are six palaeoclimate scenarios corresponding to the following geological periods, first used in Meseguer et al. (2015): Turonian (Late Cretaceous), Early Eocene, Late Eocene, Mid Miocene, Late Miocene, Mid Pliocene (Beerling et al. 2009, 2011, 2012). There are also more recent projections for the Last inter-glacial (Otto-Bliesner et al. 2008), the Last glacial maximum ($PMIP_2$, accessible from www.worldclim.org/past), and the Mid-Holocene ($PMIP_2$, accessible from www.worldclim.org/past). As observed, the number of paleoclimate projections available to the researcher increases as one gets closer to the Preindustrial age. This limitation, in particular for palaeoclimate data, implies that most ENMs have only been applied to recent geological times (Martínez-Meyer et al. 2004, Mairal et al. 2018; but see Meseguer et al. 2015). Secondly, the environmental variables available from the palaeoclimate and future climate scenarios are limited to monthly precipitation and mean temperature data (in contrast with the 7 climate and 19 bioclimate variables available for present-day projections available from WorldClim Version2, https://worldclim.org/version2). This means that one must make the assumption that these two variables alone contain enough information to describe habitat suitability satisfactorily when hindcasting ENM projections. Thirdly, and very importantly, the ENM framework implicitly assumes that species climatic tolerances have been conserved over

evolutionary time, which might be unrealistic under repeated cycles of climatic change and long geological time periods (Peterson 2006). Lastly, ENMs are evolutionary uninformative. If an entire genus is hindcasted, it is not possible to know which locality each ancestral species is present at.

THE HOLY GRAIL: REUNITING ECOLOGY, PHYLOGENY AND BIOGEOGRAPHY INTO A SINGLE FRAMEWORK TO DISENTANGLE EVOLUTIONARY AND DISTRIBUTIONAL HISTORY

To form a mechanistic understanding of the patterns formed from persistence and adaptation, geographical range shifts, local and global extinctions, and mass extinction events remains the holy grail of historical biogeography (Judson 1994; Willig et al. 2003) and ecology. Once upon a time ecology, phylogeny and biogeography were connected together by taxonomists and naturalists. However, for over the last 40 years the research field of biogeography has been concerned with phylogenies of the organisms of interest, with ecology deemed moot (Wiley 1981; Brooks and McLennan 1991; Ree and Smith 2008; Höhna et al. 2016). The research field of phylogeny has considered ecology as an informative "flavour" to the overall model and biogeography regarded moot (e.g. Condamine et al. 2013), while the research field of ecology has ignored phylogeny entirely; though lately it has strengthened itself by including species interactions, non-climatic predictors, and/or spatial autocorrelation terms (Araújo and Luoto 2007; Dormann et al., 2007; De Marco et al., 2008; Guisan and Rahbek, 2011). In the two previous sections, above I have described how different theoretical and analytical frameworks can be used to obtain a mechanistic understanding of species distribution patterns from a phylogenetic (the birth-death framework), a biogeographic (biogeographic inference), and an ecological (ENM) point of view. It is, however, important to keep in mind that the ecology of species plays a key role in interacting with evolutionary and biogeographical processes, e.g. speciation, dispersal and extinction, in the shaping of species distribution patterns (Brooks 1990; Cracraft 1994; Sanmartín et al. 2001; Wiens and Donoghue 2004). Ecology, phylogeny and biogeography have much to offer one another (Webb 2002; Wiens and Donoghue 2004; Hawkins et al. 2005), and can complement each other's limitations. An increasing number of studies has stressed the need to join these now separate areas of research into a single theoretical and analytical framework, in order to explore the relationship between ecological and evolutionary processes (Farrell et al. 1992; Ricklefs and Schluter 1993; Latham and Ricklefs 1993; Futuyma 1998; Brown and Lomolino 1998; Webb 2002; Metcalf et al. 2014).

To reunite these three research areas, ecology, phylogenetics, and biogeography, into a single mechanistic framework could allow us to better learn from the past so that we can better predict the future. Below, I describe two different approaches that have recently been developed to tackle this challenge: integrating the biogeography inference and ENM analytical frameworks, and using macroecological ("in silico") simulation models. In Chapter 2 and Chapter 3 of the thesis, I explore these two research avenues and extend them further to accommodate for limitations in distribution and ecological data in biogeographic and ENM inference (Chapter 2), and to include phylogenetic patterns within macroecological "in silico" models (Chapter 3).

Unifying biogeographic inference and ecological niche model frameworks

As mention above, ENMs provide a geographically informed framework that is absent from biogeographic inference. Some authors have advocated integrating evidence from ENMs to calibrate biogeographic/phylogeographic inferences, through either including additional areas with suitable habitat in which a taxon has become extinct or has yet to be discovered, or by postulating climatic corridors connecting areas across unsuitable habitats to evaluate dispersal or migration models (Richards et al. 2007; Smith and Donoghue 2010; Metcalf et al. 2014; Meseguer et al. 2015). These studies, in their own different ways, have attempted to reunite the three research areas: ecology, evolutionary processes and geographical distribution. Yet, rather than integrating all of them into a single mathematical framework, these studies keep the biogeographical inference and ENM frameworks separate and instead use the results to feed into one another as input data.

Although independent, these two frameworks can work well together, allowing the user to have more detailed geographical information without sacrificing evolutionary information and vice versa. Richards et al. (2007) used hindcasted ENMs and a coalescence model to create coalescent simulations and test alternative demographic hypotheses. Evans et al. (2009) used species distribution models to quantify the climatic disparity among taxa and obtain an understanding of the degree by which niches have evolved. This information was then used as secondary calibration for their dated phylogeny to predict niche occupancy through the timeline of the tree. Meseguer et al. (2015) combined fossil distributions with presence-only data to generate ENMs hindcasted over deep time. They then used this ecological information from the fossil record as secondary calibrations of habitat suitability over time in a DEC model based on extant locations, to test for the existence of potential climatic corridors or dispersal barriers specific to the lineage. Additionally, geographical information from the fossil record was an independent source of evidence to inform DEC

biogeographic inference on the presence of ancestral lineages in areas where extant species do not occur (especially important in cases of high background extinction). Meseguer et al. (2015), as well as others (Smith and Donoghue 2010; Metcalf et al. 2014), demonstrated that integrating information from the fossil record with ENMs can increase the reliability of biogeographic reconstructions (Meseguer et al. 2015) because it can add areas outside the current distribution of species, and thus escape from the constriction of using a priori defined areas.

Although the "integrative" approach described above seems very attractive, it presents several shortcomings, which makes it not appropriate for every case study. Firstly, fossil record information is not available for many organism lineages. Secondly, such level of integration of ENMs into biogeographic inference require the use of large phylogenies with many diversification events and abundance of fossil and extant occurrence data to generate reliable ENM projections over time. Lastly, while combining biogeographic inference and ENM frameworks as mentioned above provides a more accurate understanding of past evolutionary events and changes in geographic ranges over time, these methods do not allow forecasting upon future scenarios of climate change, for example, for investigating the invasion potential of a species on different biogeographic regions (Peterson et al. 2011). In Chapter 2, I propose an alternative way to overcome the limitations of combining biogeographic inference and ENM frameworks mentioned above, by combining them as independent sources of evidence that complement each other for reconstructing the evolutionary history of a depauperate, ancient species-poor lineage, the Rand Flora genus *Camptoloma* (Scrophulariaceae; see below) over multiple palaeoclimate layers, and predicting the response to future climatic scenarios.

Macroecology simulation models

In its simplest form, macroecology simulation models are mechanistic models that govern a species in a heterogeneous landscape (represented as a gridded domain) by a set of rules for location; probability and mechanism of speciation; inheritance of niche characteristics by each new species from its immediate ancestor; and conditions of the grid cell that each species must satisfy to achieve successful dispersal and colonisation (Gotelli et al. 2009). Commonly, these models are probabilistic and stochastic (an exception would be Hassell et al. 1991), and are run for a given number of time steps t or until a specific condition is met (e.g. until a particular number of species ranges are simulated or until a balance between speciation and background extinction is achieved). These models can accommodate for present day, past or future climates, evolutionary and historical forces,

and geometric constraints. These models build frameworks to investigate hypotheses on the relative role played by geometric constraints in the landscape (area size and shape, latitude, longitude, etc), climatic factors and historical processes on the shaping of species richness patterns (Gotelli et al. 2009). Moreover, they can accommodate for present day, past or future climates, evolutionary and historical forces, and geometric constraints, making them more than suitable for exploring the stability of an empirical observed relationship between diversity patterns and environmental variation. Gotelli et al. (2009) presents an excellent review of macroecology simulation models.

Understanding scenarios of persistence or geographical range shifts through simulation models has been well explored through a variety of studies. Levins (1969)'s model, which is widely considered as the null hypothesis model in macroecology simulations, was one of the first models to explore the introduction of spatial parameters into population models (Gotelli et al. 2009). Levins' population model is described as an unstable system, present in one cell of a gridded landscape, where the population would become extinct within a specified number of time steps. However, when several of these population models are linked to one another in the gridded landscape through a spatial parameter representing dispersal, extinction of the population model would only be temporary, as it could be recolonised by still occupied, neighbouring population models, hence forming a stable "on our own we are weak, together we are strong" system when population models' extinctions are in a state of asynchrony. This combined model was simple and limited itself to noting only whether a population model was occupied or not, i.e. population number was disregarded, and dispersal was considered to be unlimited to any population model within the landscape. This type of linked populations became known as the classic or Levins' metapopulation model. Further advances from these first models have included the spreading dye model, which imposed limitations on how a species may disperse through the landscape from its original cell (e.g. Hassell et al. 1991; Jetz and Rahbek 2001; Connolly 2005). GSMs have, in a way, mechanised ENMs (Thuiller et al. 2008), as including habitat suitability information in each cell has allowed for background extinction rates of cells to move from being fixed to determined by relationships between species (interspecific competition; red queen, Benton 2009; Ezard et al. 2011) and within species (intraspecific competition), and by the biotic suitability of the cell (ENM). GSMs explore the relationship between temporary barriers and the spatial distribution of a population (stressful or beneficial), and can be used to make future predictions in the changes in population distribution or numbers to explore persistence or geographical range shifts in species. In this form, however, these models can only be used to explore spatial distributions, and not evolutionary adaptation. It is possible to hindcast them

temporarily with the assumption that the relationship to temporary barriers remains unchanged (Pearson et al. 2006).

Including evolutionary events to macroecology simulation models has been tackled in previous works mainly in one of two ways: firstly, by specifying the number of independent evolutionary origin events that must take place within the simulation (e.g. Grytnes 2003; Storch et al. 2006; Rahbek et al. 2007); and secondly by including some sort of scale and logic measurement of what a species is, running the simulation for a set time and then observing how many evolutionary events occurred and how many species are present at the end of the simulation (e.g. Boone 2010 and Rangel et al. 2007). In this latter approach, evolutionary events have a single origin and depend on the algorithm that determines the balance between niche conservation and niche evolution, allowing evaluation on whether species richness on large geographical scales can be explained by this balance (Rangel et al. 2007).

To think of evolutionary events as specified independent evolutionary origin events follows the standard ecological approach to attempt to understand "why" a species richness pattern looks the way it does by looking at the absolute present day species numbers (Jetz and Rabnek 2001; Wiens and Donoghue 2004). However, this makes it difficult to compare the end result with the empirical scenario that is of interest and hence answer the question of "who did what, where and when?", which leads to the answer to "why?" (Wiens and Donoghue 2004). To decide upon a scale and logic measurement for an evolutionary event can be tricky, as the initial empirical ecological niche, the empirical heritability of the ecological niche from ancestor to descendant taxa, and an understanding of the balance between empirical niche conservation and empirical niche evolution through time must be known beforehand. One way to deal with this issue would be to follow in the footsteps of studies such Rangel et al. (2007), and remove oneself from the real world, hence alleviating the need to use empirical data. However, this removal from the real world makes it hard to study the stability and statistical robustness of an empirical, observed relationship. From the results of a virtual world simulation, one can test whether a pattern of species richness distribution similar to the observed one could be reached (Rangel et al. 2007); however, it is not possible to investigate whether a similar evolutionary relationship among species in the form of a phylogenetic tree could be recreated.

In Chapter 3, I explore a novel way to introduce evolutionary events to macroecological simulation models. Specifically, I integrate speciation and background extinction rate information from the phylogenetic tree, as well as speciation time intervals and the species evolutionary relationships represented in the cladogram.

In Chapter 3, I explore a novel way to introduce evolutionary events to macroecology simulation models. Here, I add information on speciation and background extinction rates from the phylogenetic tree, as well as speciation time intervals and the evolutionary relationships in cladogram.

CASE STUDIES

In the following section, I describe in more detail the organism lineages used as analytical case studies in the three chapters of my thesis.

Case study 1: Rand Flora genus Camptoloma

The Rand Flora (RF) pattern is a continental-scale geographic disjunction in which sister species are distributed on opposite sides of the African continent (Sanmartín et al. 2010). The classical example is the east-west disjunction between Macaronesia and Eastern Africa observed in genus *Canarina* (Campanulaceae, Mairal et al. 2015) (Figure 4). Molecular estimates have dated RF disjunctions from the Early Miocene to the Pleistocene, though most trace back to the Mid-Late Miocene and Pliocene periods, concurrent with the intensification of the aridity trend (Pokorny et al. 2015). RF lineages belong to different angiosperm families, with dissimilar morphology and life forms, but they typically exhibit subtropical, temperate or semi-arid affinities, i.e. they do not occur in deserts (Sahara) or in the tropical lowlands (Guinea-Congo Basin). Species populations are often small in size and have highly restricted distributions (Pokorny et al. 2015; Mairal et al. 2017, 2018). As in many African plant groups, there is little data on their reproductive biology or evolutionary history, and distributional records are scarce (see online occurrence databases e.g. www.gbif.org), especially for difficult-to-access or politically unstable regions (e.g. Somalia), though there have been recent efforts to address this (Mairal et al. 2017; Villaverde et al. 2018). Moreover, with a few exceptions (e.g. Thiv et al. 2010), RF phylogenies comprise less than 5-10 species and often exhibit "broom-and-handle" shapes, with long stem-branches and young crown clades, indicative of high background extinction rates or mass extinction events (Antonelli and Sanmartín 2011; Mairal et al. 2015; Sanmartín and Meseguer 2016).

One of the most extreme examples of the RF disjunction is genus *Camptoloma* (Scrophulariaceae, with three isolated species in Macaronesia (*C. canariensis* in Gran Canaria), Eastern Africa (*C. lyperiiflorum* occurring in Somalia, Yemen, Oman and the Socotra Archipelago),

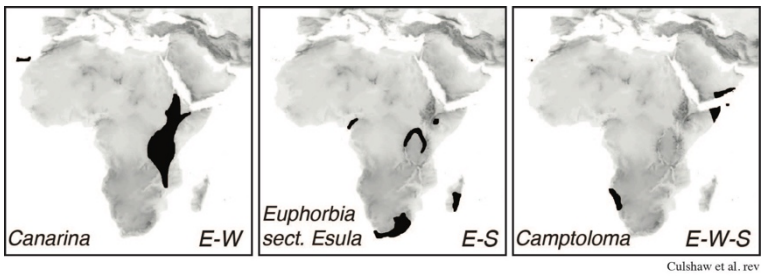
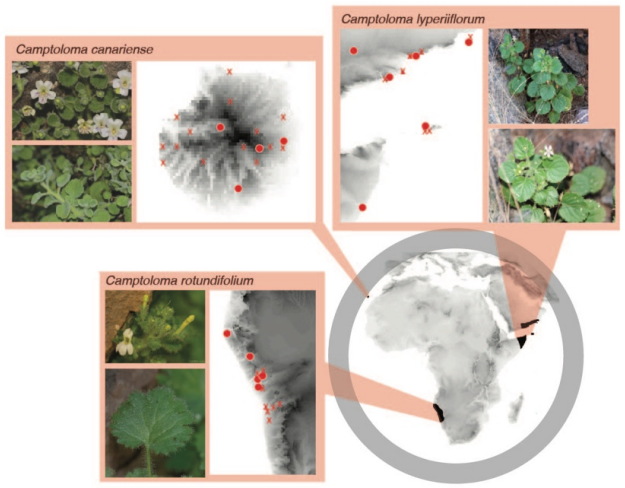


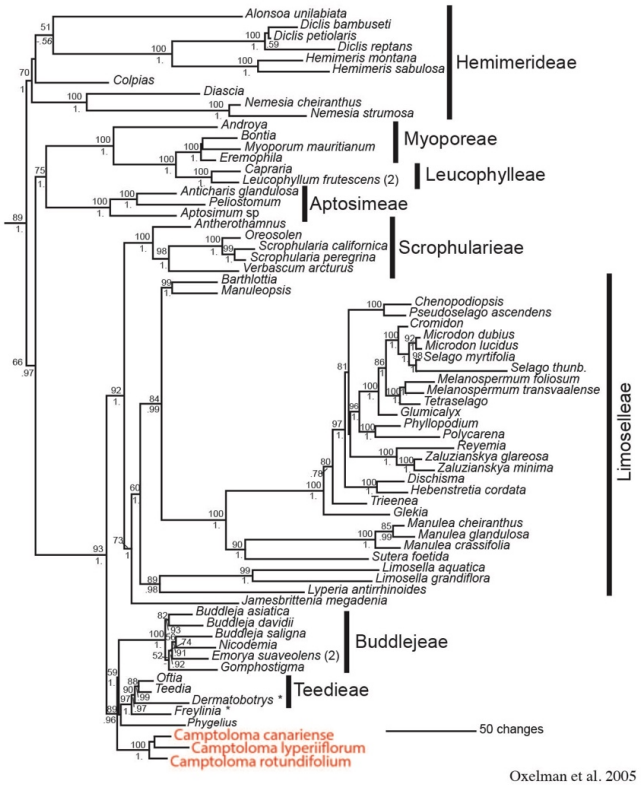
Figure 4. Rand Flora-like distributions with representative lineages. The African Rand Flora (RF) pattern is a continental-scale geographic disjunction in which sister species are distributed on opposite sides of the African continent (Sanmartín et al. 2010). The origin of the Rand Flora pattern was addressed in Chapter 2, and from a theoretically viewpoint, in Chapter 3.

and southwest Africa (*C. rotundifolium* in Namibia), shown in Figure 5). Species populations occupy climatic microrefugia, with *C. lyperiiflorum* occupying areas of humidity and low seasonality within a more arid geographical template (Domina et al. 2012), *C. canariense* in shaded vertical crevices, or *C. rotundifolium* situation in rocky crevices in the Namibian coastal Brandberg mountain range (Craven and Craven 2000).

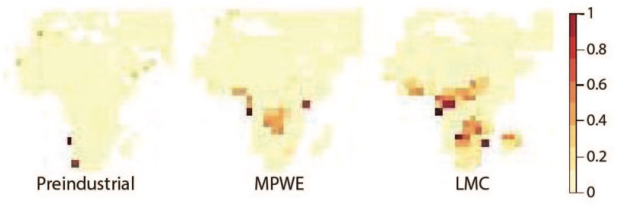
Phylogenetic relationships among species and allies in the large angiosperm family *Scrophulariaceae* remain controversial due to low phylogenetic support and limited sampling; i.e. existing molecular phylogenies have been based on three chloroplast DNA (cpDNA) markers (*ndhF*, *trnLF*, and *rps16*), with one specimen species representation (Figure 5; Kornhall et al. 2001; Oxelman et al. 2005; Pokorny et al. 2015). Genus *Camptoloma* was initially included in tribe Manuleae based on floral morphology (Hilliard 1994). In the first molecular phylogenetic analysis of this tribe, Kornhall et al. (2001) excluded *Camptoloma* and synonymised Manuleae with tribe Limoselleae. Oxelman et al. (2005) supported this classification and placed *Camptoloma* as sister to the mainly southern African tribes *Teedieae* (10 species) and *Buddlejeae* (c. 100 species; Figure 5). The three form a clade with tribe *Scrophularieae* and an extended tribe *Limoselleae* (Figure 5; Oxelman et al. 2005). A recent phylogenomic study on Buddlejeae supports this arrangement (Chau et al. 2018). No known fossils exist, meaning that dating *Camptoloma* must be done through secondary calibrations from studies such as Magallón et al. (2015). Pokorny et al. (2015) dated the origin of RF disjunction of *Camptoloma* (crown-node *Camptoloma*) in the Miocene-Pliocene boundary (5.5 Ma), whereas the divergence with its sister group (South African *Phygellius capensis*) was estimated much older, in the Early Miocene (21 Ma).



Chapter 2



Oxelman et al. 2005



Mairal et al. 2017

Figure 5. Geographic and evolutionary patterns in the RF genus *Camptoloma*: Top: Geographic distribution of extant *Camptoloma* species. Centre: Molecular phylogeny of family

Figure 5. cont. Scrophulariaceae (Oxelman et al. 2005), showing systematic placement of genus *Camptoloma*. Bottom: Geographical projection of the climatic fundamental niche of genus *Camptoloma* over three different palaeoclimate simulations representing major warming or cooling events in the Late Cenozoic Earth history (Beerling et al. 2009; Meseguer et al. 2015). These three different types of evidence were used in Chapter 2 to understand the origin of the RF distribution of genus *Camptoloma*. Abbreviations: MPWE (Mid-Pliocene Warming Event), LMC (Late Miocene Cooling event).

Representation of *Camptoloma* records on online databases and herbaria is also limited and likely biased due to the species occurrence in difficult-to-access or politically unstable regions (e.g. GBIF contains 35 records of *C. lyperiiflorum* in Yemen, 7 in Oman but none in Somalia; (9th July 2014) GBIF Occurrence Download https://doi.org/10.15468/dl.abs9dk). This is problematic when using statistical methods for integrating ENMs and biogeographic inference, as has been done in other RF lineages with larger fossil occurrence datasets (Meseguer et al. 2015, 2018). Mairal et al. 2017 presents projections of *Camptoloma*'s geographic fundamental climatic niche created by an ENM over three palaeoclimatic layers representing major warming or cooling events in the Late Cenozoic Earth history (Figure 5). These projections indicated a general decrease in climatic suitability through time since the mid-Miocene (Figure 5). The projections in the Late Miocene period, representing a global cooling event, revealed wide distribution ranges with connections joining the east and west edges of Africa with an environmentally suitable corridor south of the Sahel. These projections, however, provide only a limited insight as to how *Camptoloma* responded to climate change over millions of years. Due to the small-sized phylogeny (3 tips), large temporal gaps, limited species occurrence data and the lack of fossil record in RF lineages (Sanmartín et al. 2010), it is difficult to use sophisticate diversification and ancestral reconstruction methods for inferring evolutionary history (Sanmartín and Meseguer 2016).

The evidence that these three species have been isolated for several million years (Pokorny et al. 2015) is in agreement with the hypothesis of climatic relicts. These deep Miocene divergences, broom-and-handle-shaped phylogeny, and microrefugia-type distributions makes *Camptoloma* an ideal case study to test the hypothesis of RF disjunctions as climatic relicts linked to African aridification history, and to explore their potential response under scenarios of future global climate warming.

Case study 2: Conifer phylogeny (*Leslie* et al. *2012)*

Leslie et al. (2012) presented a fossil-dated phylogeny of conifers, containing 492 of 630 known species (78%; Mabberley 2008). It is divided into two clades: northern and southern, and has an inferred crown age of 340-343 Ma (Figure 6). This phylogeny describes the relationship between and within the clades located in the Northern and Southern Hemisphere, with the initial divergence within each clade starting at 328.16 Ma. Leslie et al. (2012) demonstrated that most extant conifer species diverged in the Neogene, but belong to major lineages that generally diverged earlier in the Mesozoic. Locations of extinct conifer species are well represented in a massive fossil record, which dates back to the Pennsylvanian (323.2-298.9 Ma; Scott and Chaloner 1983; Plotnik et al. 2009).

Due to the plethora of data available for both extant and extinct lineages, the conifers have been intensely studied, from the characterisation of their reproductive morphology (Tomlinson et al. 1991; Takaso and Owens 1995; Owens et al. 1997; Tomlinson and Takaso 2002; Farjon and Ortiz Garcia 2005), to their pollination biology (reviewed in Owens et al. 1998) and seed dispersal biology (Benkman 1995; Siepielski and Benkman 2007a, b, 2008). Höhna et al. (2015) and May et al. (2016) used Leslie et al. (2012)'s phylogeny to test their methods for detecting mass extinction events: they located a significant MEE around 23 Ma that coincides with the Late Oligocene Warming Event, and two potential mass extinction events occurring around 85 Ma and 173 Ma (Figure 1j, 6). Using BEAST dating, Leslie et al. (2012) demonstrated that the turnover for the northern hemisphere lineages was higher than for the southern hemisphere lineages.

Leslie et al. (2012)'s study indicated that there are large-scale and consistent differences between the diversification of southern and northern conifer clades, with the evolutionary dynamics of the conifer lineages inhabiting mainly southern hemisphere environments, including both temperate and tropical habitats, differing from the evolutionary dynamics of Northern Hemisphere lineages. Specifically, northern hemisphere lineages have a greater proportion of very recent divergence times (i.e. within the past 5 My) and fewer deep divergences than southern hemisphere lineages. These general differences in climatic and landscape history, resulting from the dispositions of landmasses in the hemispheres, appear to have left a distinct imprint on conifer evolutionary history.

Recently, Wang and Ran (2014) reconstructed the evolutionary and biogeographic history of gymnosperms, exploring the roles played by vicariance and dispersal in the shaping of intercontinental disjunctions (Figure 6). They included the conifer species in Leslie et al. (2012)'s phylogeny, and their biogeographic analysis suggested that most conifer genera originated in North

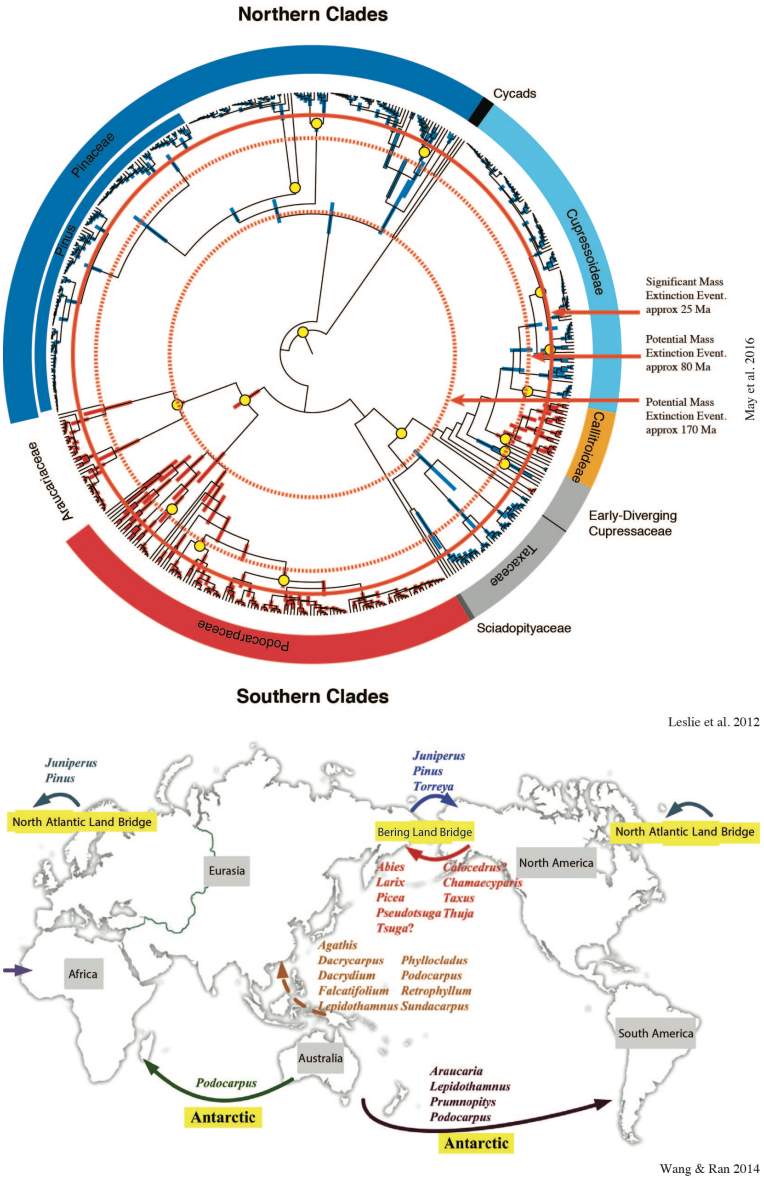


Figure 6. Phylogenetic relationships of the conifers (subclass Pinidae), including families Cupressaceae, Pinaceae, Araucariaceae, etc (Leslie et al. 2012). A dated phylogeny for 489 conifer species with cycads as an outgroup, shows that most extant conifer species diverged in the Neogene, but belong to major lineages that generally diverged much earlier in the Mesozoic. Potential (red dashed line circles) and significant (solid red line circle) mass extinction events. Inferred dispersal route of the intercontinentally dispersed genera. Question marks indicate migration routes of taxa have not been resolved, while dash arrow indicates a possible of migration. This megaphylogeny was used for statistically testing the power of the Birth-Death Skyline model for inferring MEEs explored in Chapter 1; it was also used as a baseline example for testing the in-silico model in Chapter 3.

America and migrated into the Old World, thus supporting the contention that the centre of diversity of a group is not necessarily its place of origin. Wang and Ran (2014)'s biogeographic reconstruction is valuable, as it provides hypotheses on where different conifer lineages originated. For example,

Wang and Ran (2014) suggested that *Pinus* has an Eurasian origin and underwent multiple dispersal events between Eurasia and North America via the North Atlantic Land Bridge (NALB) (Figure 6; Eckert and Hall 2006; Mao et al. 2010)

The conifers make an ideal study case to explore the signature of mass extinction events (i.e. the MEEs inferred by Höhna et al. (2015) and May et al. (2016)), and the role of macroecology and global scale biogeography in triggering and modulating these events. Such knowledge may help better pinpoint the timing and location of the mass extinction events, and to explore methods to further learn on the diversification history and dispersal routes of specific conifer genera (e.g. *Pinus;* Figure 6).

AIMS, HYPOTHESES AND OBJECTIVES

OVERARCHING AIM

The overarching aim of this thesis is to develop new tools and methodologies to understand the evolutionary and distributional signature of plant lineages' responses to the impact of climate change. These responses are defined as: i. adaptation of the species that can potentially lead to speciation; ii. persistence at the geographical location; iii. migration from the geographical location in the form of displacement, expansion and contraction; iv. local extinction of a population leading to fragmentation of the geographic distribution and/or global extinction of the species; and v. mass extinction of lineage(s). In order to achieve this overarching aim, we will employ four different frameworks: birth-death models, biogeographic inference methods, hindcasted and forecasted ecological niche models, and macroecology simulation models. Also, we consult two case studies: Leslie et al. 2012's conifer phylogeny that contains a plethora of information; and depauperate Rand Flora genus *Camptoloma*. Within this thesis, solutions to deal with poverty rich data will be explored, with poverty rich being in the form of: phylogenetic trees composed of extant taxa only; phylogenetic trees with few tips; and spatial and temporal limited palaeoclimate and future projected data.

CHAPTER 1 AIMS: EXPLORING THE POWER OF BAYESIAN BIRTH-DEATH SKYLINE MODELS TO DETECT MASS EXTINCTION EVENTS FROM PHYLOGENIES WITH ONLY EXTANT TAXA

In this chapter we have three general aims: first, we explore the power of the birth-death model, "Bayesian birth-death skyline" (BDSKY; Stadler et al. 2013), as a new approach in identifying and locating instantaneous mass extinction events (MEE) through changes in the diversification rate in simulated extant taxa only phylogenetic trees; second, we compare BDSKY's performance to that of the "Compound Poisson Process (CPP) on Mass Extinction Times" (CoMET; May et al. 2016), a model that was specifically designed to read MEE signatures; and third, we compare the results of the BDSKY model to that of the CoMET in identifying MEE in Leslie et al. 2012's conifer phylogeny. The underlying hypothesis is that the signature of an instantaneous MEE can be recovered in phylogenetic trees by BDSKY just as well as CoMET. The particular tasks are:

- Create a simulation study of extant taxa only phylogenies that are simulated under the single-pulse model, and have been affected by one MEE. These trees differ from one another by size, background extinction rate, MEE time and mass extinction intensity.
- Assess the power of BDSKY and CoMET to detect and locate the MEE signature in these simulated phylogenies and estimate speciation and background extinction rates pre- and post-MEE of these phylogenies.
- Compare the performances of BDSKY and CoMET to the simulated phylogenetic tree in terms of their frequency of Type I and Type II errors, and their accuracy in estimating the magnitude of speciation and extinction rates and the timing of MEEs.
- Apply BDSKY to Leslie et al. 2012's conifer phylogeny.
- Compare the results of BDSKY to CoMET to look for the signature of MEE(s) in Leslie et al. 2012's conifer phylogeny.

CHAPTER 2 AIMS: COMBINING BAYESIAN BIOGEOGRAPHIC INFERENCE AND PHYLOGENETICALLY-INFORMED NICHE MODELS TO RECONSTRUCT THE ROLE OF ANCIENT CLIMATE CHANGE IN DEPAUPERATE LINEAGES

The two aims of this chapter are to: first, present a novel methodology for rejoining ecology, phylogeny and biogeography in order to disentangle historical biogeography by using results from environmental niche models (ENM) and Bayesian biogeographic inferences in tandem, hence keeping these frameworks as independent but complementary sources of evidence, to infer the spatiotemporal evolution of species poor lineages with restricted distributions and limited associated data. Here we employ as case study Rand Flora's *Camptoloma* and test the hypothesis of Rand Flora disjunctions as climatic relicts linked to African aridification history and *Camptoloma*'s survival under scenarios of global climate warming; and second, present a novel truncation threshold methodology, which is informed by phylogenetic biogeographic inference and is more resistant to biased and limited data, in order to transform the continuous predictions of ENMs into presence/absence for defining habitat suitability in ENM models for past and future climate change scenarios. The underlying hypothesis is that it is possible to infer the influence climate change has had upon the evolutionary and distributional history of the depauperate genus *Camptoloma,* by drawing upon different sources of evidence that complement one another with respect to their data limitations. Keeping the ENM and Bayesian biogeographic frameworks separate, stems from the logic that if two separate frameworks, which both

draw upon different pools of information, are able to return similar results, then these models testify for one another's reliability. The particular tasks are:

- Resolve and date phylogenetic relationships among the three species of genus *Camptoloma* at the population level and their systematic position within Scrophulariaceae at the species level using seven noncoding cpDNA markers and the nuclear ribosomal ITS region and secondary fossil calibrations.
- Reconstruct ancestral ranges and the history of geographic range shifts of Scrophulariaceae and the three species of genus *Camptoloma* using hierarchical Bayesian inference approaches that are appropriate at a population and species level.
- Create ENMs of the genus *Camptoloma* and each of its species.
- Hindcast the ENMs of the genus *Camptoloma* upon palaeoclimatic layers representing global warming or cooling events from the Early Miocene to the present day to predict changes in the ancestral species geographic ranges over time.
- Combine information from Bayesian biogeographic inference and palaeoclimatic hindcasted ENMs to infer
- the evolutionary and distributional history of the genus *Camptoloma* under climate change.
- Forecast the fate of the genus *Camptoloma* under scenarios of future climate warming.
- Demonstrate that biogeographic inferences and ENMs can be used to together to verify lineage dating and ancestral fundamental and realised state niche reconstruction even though palaeoclimate data is temporally and spatially limited, no known fossils exist, and phylogenies are small with sparse branches and no extinct taxa.

CHAPTER 3 AIMS: EVOLUTIONARY MACROECOLOGY: INCORPORATING PHYLOGENETIC INFORMATION MORE EXPLICITLY INTO PROCESS-BASED, FORWARD TIME, SIMULATION MODELS

In this chapter, we aim to present a novel single origin, evolutionary component to a process-based, pattern orientated, forward time, evolutionary macroecology simulation model that assumes phylogenetic niche conservation. The evolutionary component is informed by: a speciation rate (λ), to determine speciation event occurrences; a background extinction rate (μ), influenced by local abiotic and biotic dynamics, where the accumulation of local extinctions ultimately results in a global extinction; and the further options of adding the tree cladogram and/or speciation time intervals. All information within this evolutionary component should be derived from an empirical or simulated

dated phylogenetic tree. The main two advantages of this evolutionary component are: first, the simulation model's results can be more directly compared to the empirical or simulated example; and second, if there is a lacking of estimations or reliability of λ and μ, then information from the cladogram and speciation time intervals can be used to explore the λ and μ space to find appropriate pairings. Unlike the ecological niche model and biogeographic inference methodology present in Chapter 2, which provides precise and rapid insights of the evolutionary and distributional history of an empirical or simulated example, the underlying hypothesis is that this model will be able to explore the stability limits of the relationship between evolutionary and distributional history i.e., is the phylogentic tree directly linked to the distribution patterning we see or could the same phylogentic tree have been achieved with different distribution patternings, and vice versa? Knowing the stability limits of the relationship gives a better understanding of the model's power of predicting future evolutionary events and distributional patterns of the empirical or simulated example as it reacts to changes in its environment, whether those changes come from, e.g., future predicted climate change or man made changes to the landscape such as a hydroelectric dam. The particular tasks are:

- Present and code the model with its novel single origin, evolutionary component.
- Present a methodology to use the model for the two case studies: the depauperate *Camptoloma* genus, which is too small to have reliable λ and μ estimations from a birth-death model; and the *Pinus* genus from Leslie et al. 2012's conifer phylogeny.

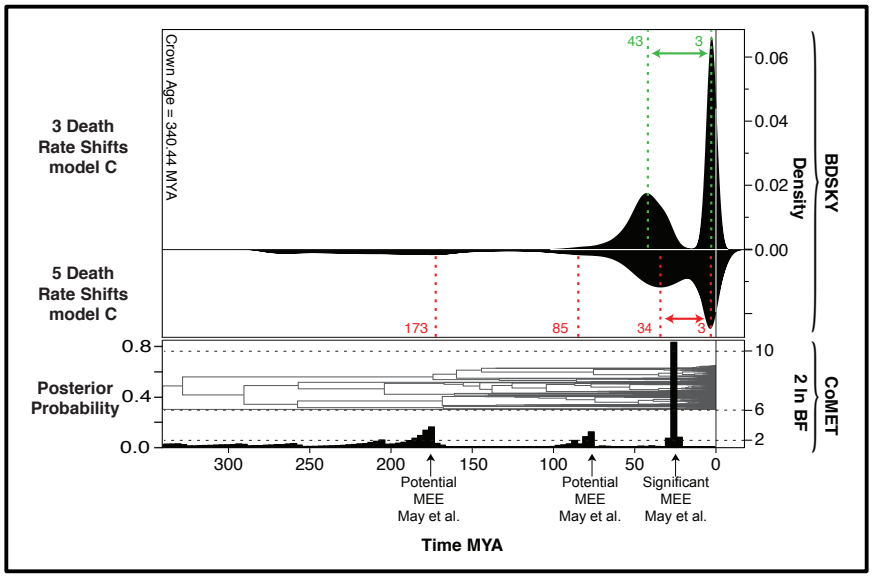
CHAPTERS

CHAPTER 1

Exploring the power of Bayesian birth-death skyline models to detect mass extinction events from phylogenies with only extant taxa

This chapter has been realised in collaboration with Tanja Stadler (computational evolution group, Department of Biosystems Science and Engineering, Eidgenössische Technische Hochschule Zürich, Basel) and Isabel Sanmartín (Department of Biodiversity and Conservation, Real Jardín Botánico, CSIC, Spain).

This chapter has been published in the peer-review journal *Evolution*, volume 73 (6) 1133-1150 (2019)



CHAPTER 1

Exploring the power of Bayesian Birth-Death Skyline models to detect mass extinction events from phylogenies with only extant taxa

Victoria Culshaw, Tanja Stadler, Isabel Sanmartín

ABSTRACT

Mass extinction events (MEEs), defined as significant losses of species diversity in significantly short time periods, have attracted the attention of biologists because of their link to major environmental change. MEEs have traditionally been studied through the fossil record, but the development of birth-death models has made it possible to detect their signature based on extant-taxa phylogenies. Most birth-death models consider MEEs as instantaneous events where a high proportion of species are simultaneously removed from the tree ("single pulse" approach), in contrast to the paleontological record, where MEEs have a time-duration. Here, we explore the power of a Bayesian Birth-Death Skyline (BDSKY) model to detect the signature of MEEs through changes in extinction rates under a "time-slice" approach. In this approach, MEEs are time intervals where the extinction rate is greater than the speciation rate. Results showed BDSKY can detect and locate MEEs but that precision and accuracy depend on phylogenies size and MEE intensity. Comparisons of BDSKY with the single-pulse Bayesian model, CoMET, showed a similar frequency of Type II error and neither model exhibited Type I error. However, while CoMET performed better in detecting and locating MEEs for smaller phylogenies, BDSKY showed higher accuracy in estimating extinction and speciation rates.

Running-title: Mass extinction birth-death modeling

Keywords: Bayesian skyline birth-death model, diversification rates, episodic models, extinction, mass extinction events, speciation

INTRODUCTION

Mass extinction events (MEE) are distinguished in the paleontological records as widespread, higher taxonomic group extinctions (e.g. up to 96% of marine invertebrate species became extinct during the largest, late Triassic MEE; Raup 1979). A MEE is defined as a period where (i) the ratio of the extinction rate μ over the speciation rate λ, aka the turnover rate, $\varepsilon = \frac{\mu}{\lambda}$, is larger than 1; and (ii) this period is less than one million up to about 15 million years (Mya) duration, dependent on the magnitude or intensity of the MEE (Sepkoski 1982). This often results in an ecosystem's speedy decline and reordering (Gould 1994). MEEs are monocyclic (irregularly cycling) and are generally recognized as a resultant of abiotic changes (Sepkoski 1982; Jablonski 2008), which are often compared to current-day greenhouse-induced climate change (IPCC 2001). MEEs have been usually studied from paleontological evidence (Raup 1979; Sepkoski 1982; Jablonski 2008). High extinction rates, as those associated with MEEs, can also leave an imprint on the timing and structuring of

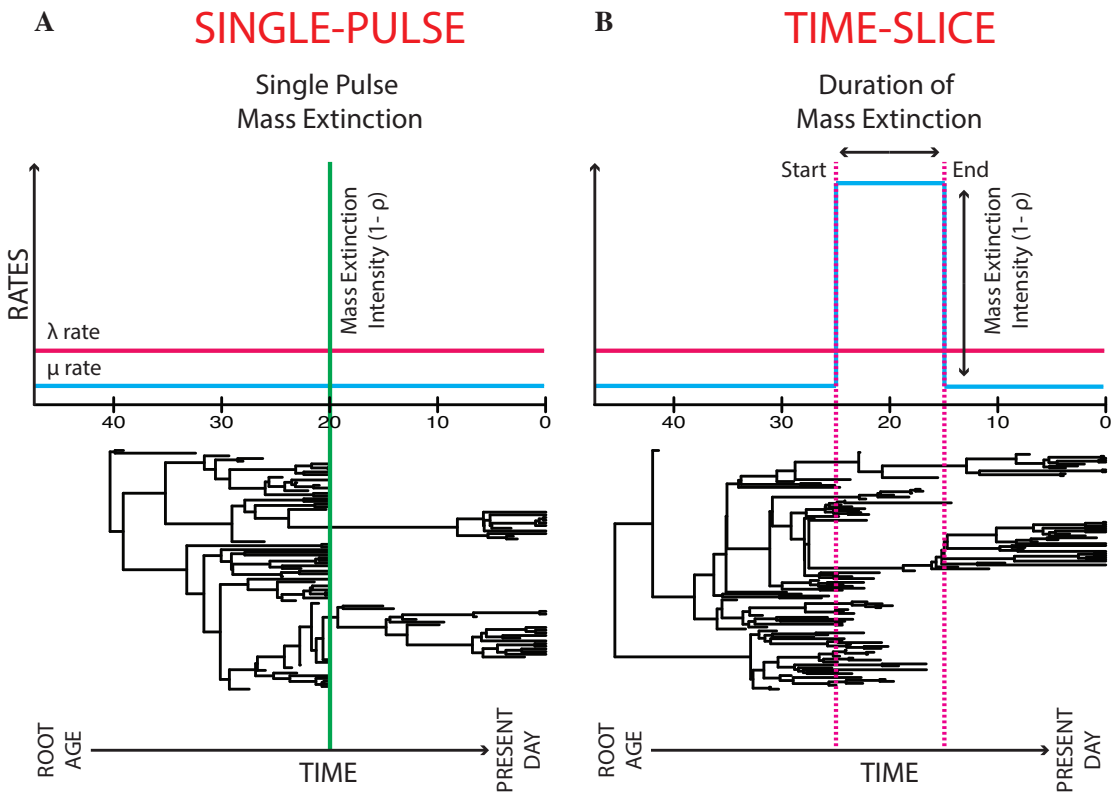


Figure 1: Two examples of full (extant and extinct taxa) phylogenetic trees that contain 20 taxa at time *t*=0 and have similar root ages. The first tree has been affected by a MEE that is defined under the "single-pulse" scenario, and the second tree has been affected by a MEE, defined under the "time-slice" scenario. Within these scenarios speciation rate, λ is assumed to be unchanged. In the "single-pulse" scenario the MEE is caused by a significant percentage of species being simultaneously and instantaneously removed from the tree, at a specified time. In the "time-slice" scenario, the MEE is defined as a significant increase in the extinction rate, μ for a specific period of time, where the turnover rate, $\varepsilon = \frac{\mu}{\lambda} > 1$, followed with a decrease in μ that results in a return to $\varepsilon < 1$. In the two trees, the "pre-MEE" μ is equal to "post-MEE" μ but this is not necessary.

cladogenetic events in phylogenetic trees containing only extant taxa (Harvey *et al.* 1994). This has permitted a burgeoning of method development in the context of macroevolutionary birth-death models for detecting the phylogenetic signature of MEEs. Methodologically –within the birth-death model framework– an MEE detecting model can be defined as either: (i) a "single-pulse" model, in which a significant number or percentage of species are instantaneously and simultaneously removed from the phylogenetic tree at a specific point in time *t* (Stadler 2011c; Figure 1a); or (ii) a "time-slice" model, in which there is a significant increase in the extinction rate μ for a short time interval, with turnover rate $\varepsilon > 1$, followed by a decrease in μ that returns $\varepsilon < 1$ in the next time interval (Figure 1b). The first definition, the single-pulse model, is the one used more often in mathematical birth-death modeling (Harvey *et*

al. 1994; Stadler 2011b; Laurent *et al.* 2015; May *et al.* 2016).

If a clade is associated with a sufficient fossil record, then this record can be used to quantify the most probable number of MEEs within the specific time period that spans the fossil data (Raup, 1979). However, most clades either possess an incomplete fossil record or lack a fossil record entirely. Phylogenetic trees of extant species can also be incomplete as a result of incomplete extant species sampling, roughly defined as either: (i) a random fraction of missing species in the phylogeny; or (ii) clades in the phylogeny are collapsed to single tips due to higher taxonomic sampling level. Constant-rate birth-death models can accommodate for such incomplete extant taxon sampling through the introduction of a sampling parameter (Höhna *et al*. 2011; Stadler and Bokma 2012). These constant-rate birth-death models have been expanded to estimate speciation and extinction rate shifts through time and search for the presence of possible MEEs in the past (Stadler 2011a; Höhna, 2014).

Stadler (2011b) introduced a birth-death model that is able to detect the presence of rate shifts based on phylogenetic trees containing extant taxa only. This model assumes discrete time intervals during which the speciation and extinction rate is constant, and the rates may change arbitrary between intervals (Stadler 2011b). Such changes may be due to time-slice MEEs or other non-mass-extinction rate changes. The model can also be used to detect instantaneous MEEs following the single-pulse model defined above –points in time in which the standing diversity is reduced by a significant fraction that is controlled by the magnitude of the MEE, with the magnitude being defined as 1 minus the survival probability of each species at the MEE (ρ). Stadler's birth-death model was implemented within a maximum likelihood (ML) framework in the R package *TreePar* (Stadler 2011c), and used successfully to detect the timing of MEEs in phylogenies that have a large to moderate number of terminals (e.g. N > 200-500 taxa; Laurent *et al.* 2015; Beaulieu and O'Meara 2015). Sanmartín and Meseguer (2016) found that this model underperforms with relatively small phylogenies (N < 50 taxa).

For phylogenetic trees that span millions of years, it is likely they have been affected by rate shifts and single-pulse MEEs, perhaps caused by global (climatic or geological) events (Laurent *et al.* 2015). However, though Stadler (2011b)'s model can be used in principle to estimate the timing and magnitude of single-pulse MEEs from extant-only taxa, it remains difficult to simultaneously estimate the frequency of tree-wide rate shifts in diversification and single-pulse MEEs due to issues of parameter non-identifiability, i.e. when different combinations of parameter values yield flat likelihood surfaces for part of the parameter space (Rannala 2002). In fact, under a ML framework, it remains impossible to distinguish between a constant birth-death

process with single-pulse MEEs and a process in which diversification rates vary discretely over time because both types of processes generate identical phylogenetic signatures and have comparable likelihood functions (Stadler 2011c; Sanmartín and Meseguer 2016). Hence, in *TreePar*, one of these parameters must be fixed, for example, by assuming that μ and λ have remained constant before and after the single-pulse MEE event, or by fixing the intensity of the MEE before inferring the timing and number of rate shifts (Stadler 2011b). Also, the algorithm cannot estimate multiple rate shifts simultaneously; instead, it uses a greedy approach where the time of one rate shift is estimated and fixed before estimating the time of the next rate shift (Stadler 2011b).

To overcome this overparameterization issue, May *et al.* (2016) introduced a Bayesian statistical inference approach to the single-pulse MEE model, the Compound Poisson Process (CPP) on Mass Extinction Times (CoMET), implemented in the R package *TESS* (Höhna *et al.* 2015). Bayesian inference is less problematic under overparameterization than equivalent likelihood-based approaches due to the integration of parameter uncertainty through estimation of marginal likelihoods. CoMET implements a stochastic branching process model in which rates of speciation and extinction are constant between rate shifts, and single-pulse MEEs are modeled as tree-wide instantaneous extinction events. Specifically, the method considers three types of events: instantaneous tree-wide shifts in speciation rate, instantaneous tree-wide shifts in extinction rates, and instantaneous tree-wide single-pulse MEEs. Each of them is modeled through a separate CPP, with waiting times distributed exponentially according to event-specific rate parameters (May *et al*. 2016). To address the problem of parameter non-identifiability in single-pulse models, CoMET implements a hierarchical Bayesian approach in which rate shifts in speciation and extinction are considered as "nuisance" parameters that are integrated over in the estimation of the marginal posterior probabilities of the focal parameters: the time, number and the intensity (magnitude) of single-pulse MEEs. CPP models themselves are very sensitive to the choice of priors, which means that in practice some parameters of the model, such as the magnitude of the single-pulse MEE, are assigned informative empirical priors (May *et al*. 2016).

Here, we examine a different type of approach, the "Bayesian birth-death skyline" (BDSKY) model, first introduced by Stadler and collaborators to trace temporal changes of epidemic spread in an infectious disease (Stadler *et al*. 2013; Boskova *et al*. 2014). Stadler and collaborators (2013) used simulations to explore the power of the BDSKY model to detect changes in the rate of becoming non-infectious, akin to time-slice MEEs in an epidemiological context. In this study, we explore the power of the BDSKY

model to detect MEEs from species phylogenies under a "time-slice" (paleontological) approach, that is, through sequential changes in extinction rates, where a shift to negative diversification rates and background extinction > 1 is followed –after a (geologically) short time interval– by a return to positive diversification rates (Figure 1b; Condamine *et al.* 2013, cf Figure 4; (May et al. 2016). Unlike in Stadler *et al.* (2013)'s phylogenies, which included serial sampling (tips of different age) and were simulated under the BDSKY model, our reconstructed species phylogenies included only contemporaneous tips and mass extinction events were simulated under the single-pulse model. The aim was to assess the ability of BDSKY as a statistical phylogenetic method for detecting and estimating the timing of MEEs through changes in the extinction rate ("time-slice" approach), even when mass extinction is modeled as an instantaneous event in the same manner as in *TreePar* and CoMET.

We set up an extensive simulation study in which phylogenies were generated under a constant-rate birth-death process with one single-pulse MEE with intensity or magnitude $1 - \rho$. In these simulations, we sequentially varied different parameters to explore their influence on the power of BDSKY to detect and estimate the timing of MEEs: the survival probability of a species at the time of the MEE (ρ); the number of tips or extant taxa in the phylogeny (N); and the magnitude of μ relative to λ or background extinction (ε). We then compared the simulated values with the marginal posterior probability distributions of the model parameters estimated under the BDSKY model by Bayesian MCMC in *BEAST2*. Specifically, we aimed to answer the following questions: (i) Can we detect the phylogenetic signal of a MEE in the extant phylogenetic tree through changes in the diversification rate? (ii) If we are able to identify the presence of a MEE, can we accurately estimate the timing of that MEE? (iii) If we can identify and locate the MEE, can we provide reasonable estimates for λ and μ pre- and post-MEE?

In addition, we compared the performance of BDSKY against CoMET (May *et al*. 2016), which is implemented in the R package *TESS* (Höhna *et al.* 2015). We analyzed a subset of the simulated phylogenies above and then compared the behavior of the two models in terms of their frequency of Type I and Type II errors and their accuracy in estimating the magnitude of speciation and extinction rates and the timing of MEEs. Finally, we applied BDSKY to the conifer phylogeny of Leslie *et al.* (2012), which was also analyzed under CoMET by May *et al.* (2016), and therefore provides a test study to investigate the robustness of the two models with an empirical dataset.

CHAPTER 1

MATERIAL AND METHODS

Simulation design

The phylogenetic trees were simulated using the backward algorithm implemented by the function *sim.rateshift.taxa* from the *TreeSim* R package (Stadler 2011c). This algorithm simulates birth-death trees with speciation rate λ, extinction rate μ, a MEE at time t before the present, and survival probability of a species at the MEE, ρ. Simulated trees were conditioned on a fixed number of extant taxa (Stadler, 2011c).

All trees were simulated to include a single-pulse MEE at time t, and assuming equal and constant rates of μ and λ before and after the MEEs and across all clades in the phylogeny. Although the assumption of constant rates and rate homogeneity across clades is empirically unsound (Rabosky 2014), using this constant birth-death process provides insights into the ability to infer MEEs in the simplest scenarios, and facilitates the comparison with other studies that examined the power of episodic birth-death models for MEE estimation (Laurent *et al.* 2015; May *et al.* 2016). Since we are interested here in the power to detect MEEs through changes in extinction rates, λ was fixed to 0.2, whereas the rate of extinction was allowed to vary across simulation scenarios. One hundred trees were simulated for each of the following scenarios and (ten) parameter combinations: varying background extinction rate ε (μ = 0, 0.1, 0.18) with fixed number of extant taxa (N = 500) but under different survival probability scenarios (ρ = 0.9, 0.5, 0.1), and varying number of extant taxa (N =100, 200, 500) and background extinction rate ε (μ = 0, 0.1, 0.18) but with fixed (low) survival probability (ρ = 0.1).

We considered that all present-day species were included in the phylogeny (i.e. taxon sampling was complete at present). The MEE was simulated under a "field of bullets" scenario, where all taxa have the same probability of becoming extinct ($1 - \rho$), which is often considered as the null model for mass extinction scenarios (Raup 1979; Harvey *et al.* 1994; Laurent *et al.* 2015). For each parameter combination, we set up a control or null model corresponding to a scenario where there is no MEE (ρ = 1 at time t); again we simulated 100 trees for each parameter combination with ρ = 1. In all, our simulation study included 2400 trees.

We selected varying values of t (the time of the single-pulse MEE) for the varying values of μ, such that we could get a comparable number of lineages at the time of the MEE in all simulations (~4000 lineages; Figure S1 and S2); in particular we chose $t = \frac{0.4}{(\lambda-\mu)}$. The final simulations were done with μ = 0.18, t = 20; μ = 0.1, t = 4; and μ = 0, t = 2.

Parameter estimation under the BDSKY model

We used the BDSKY model implemented in BEAST v2.2.1 (http://beast2.org) to estimate the marginal posterior probability distribution of the model parameters using Bayesian MCMC for all simulated trees under the scenarios described above and their corresponding controls. Some parameters in the BDSKY model were set to fixed values: the simulated trees were fixed (i.e. not estimated from sequence data); sampling before present was fixed to 0, as we have no serial sampling through time (i.e. only the extant reconstructed tree is considered); and extant species sampling was set to 1 as in the simulations. We set the number of rate shifts to two, defining three time-intervals (pre-MEE, MEE, post-MEE), and estimated the speciation and extinction rates in each time interval, as well as the two rate-shift times bounding the MEE time interval (i.e. the time interval assumed to contain the MEE). We compared this full BDSKY model, in which all free parameters are estimated, against constrained BDSKY models, where some parameters are fixed:

A. Full model: All eight parameters were estimated: the two rate-shift times bounding the MEE time interval and the values of λ and μ in the pre-MEE, MEE, and post-MEE time intervals.

B. Constrained-time model: The rate shift times were fixed to ± 5% around the MEE (i.e. $t \pm 0.05\,t$), but μ and λ were allowed to vary and estimated for each of the three time intervals: $[0, t - 0.05t), [t - 0.05t, t + 0.05t], (t + 0.05t, root\ age]$. For example, for t = 20, these time intervals would be [0, 19), [19,21], (21, root age]; for t = 4, they are [0, 3.8], [3.8,4.2], [4.2, root age)].

C. Constrained-speciation model: The time interval around the MEE was estimated (i.e., the rate-shift times bounding the MEE interval) and μ was allowed to vary across time intervals, but the speciation rate λ was estimated and assumed constant (i.e. no speciation rate shifts allowed).

Priors: We explored through initial analyses different prior distribution choices available in BEAST2. These were set eventually to the following: the extinction and speciation rates, μ and λ, were each modelled with an *exponential* distribution (*Exponential (0.25)*). The rate shift times for μ and λ in Models A and C were modelled with a *uniform* prior distribution, with the lower boundary being (root age of the tree) × 0.05 and the upper boundary equal to (root age of the tree) × 0.95. In Model B, the parameters "*birthRateChangeTimes*" and "*deathRateChangeTimes*" were assigned fixed values (see above). We used the forward-in-time approach ("*reverseTimeArrays* set to FALSE) for defining the time intervals between the two rate shifts (Stadler et al. 2013). Thus, the first time interval is the oldest interval in the phylogeny (pre-MEE); the second interval is the time interval assumed to contain the MEE,

and the third interval is the youngest time interval (post-MEE). The complete xml files used in the three analysis settings (Models A to C) are provided in Supplementary S16a. To aid in the implementation of these analyses in BEAST2, we also provide a new R function (that uses R packages *TreeSim, picante* and their dependents) that writes the xml code necessary to run a full or constrained BDSKY models for species trees, and example code for running this function; see Supplementary S16b-e.

Analysis: Each model was run to convergence, when the effective sample size (ESS) for each estimated parameter reached a value equal or larger than 200. MCMC runs that were unable to converge after 1000 million generations were discarded but not replaced, so the final number of analyzed simulated trees differed between prior settings, though this was never lower than 86% (Table S1). We compared the performance of the three BDSKY model settings (Models A to C) in terms of Type I error or the percentage of false positives – detecting the presence of a MEE when none was simulated– and Type II error or false negatives –failing to detect the simulated MEE; the latter is equivalent to 1 minus "power", i.e., the proportion of trees analyzed for which the method correctly detected the MEE. We defined the detection of a MEE as the 95% High Posterior Density (HPD) credibility interval for the diversification rate estimate ($r = \lambda - \mu$) being negative in the MEE time interval (and not containing 0), followed by a return to positive diversification rates (and potentially containing 0) in the third time interval. Hence, we defined a new statistic to measure performance: *HPDn* is the percentage of simulated trees for which all values within the 95% HPD for the diversification rate are lower than 0 (and the 95% HPD for ε is fully above 1) in the MEE interval, *and* the 95% HPD for diversification rate is fully above or containing 0 in the post-MEE time interval. We also estimated the robustness of BDSKY under each model setting in the estimation of the non-fixed parameters. For each simulation scenario (with and without a MEE), we measured: the *accuracy*, the true (simulated) parameter value minus the mean of the estimated parameter means for each tree; the *precision*, the mean of the width of the 95% HPD interval across all analyzed trees; and the *coverage*, the percentage of simulated trees where the 95% HPD contained the true parameter value. In particular, we focused on the ability of the BDSKY model to estimate the posterior distributions of the parameters: diversification rate (r), speciation rate (λ) and extinction rate (μ) in each time interval, as well as the two rate-shift times bounding the MEE time interval. For this interval, the true diversification rate can be calculated with respect to the true mass extinction intensity and the rates of λ and μ as:

$$True\ Diversification\ Rate\ in\ MEE\ interval = \lambda - \mu - \frac{1-\rho}{length\ of\ time\ interval}$$

The mass extinction rate $s = \frac{(1-\rho)}{length\ of\ time\ interval}$ summarizes the instantaneous mass extinction intensity. The rate s is obtained by recalling that the probability for mass extinction is $s \times (length\ of\ time\ interval) = (1-\rho)$ given that the time interval is short.

For those analyses in which rate shift times were not fixed but estimated (Models A and C), we do not have a corresponding true value. Instead, the MEE occurred at time *t* while we estimate a time interval $(t1, t2)$ during which the extinction rate exceeded the speciation rate, signaling the presence of the MEE. In these two models, *precision* was measured as the mean of the 95% HPD interval of the time length $(t1 - t2)$ across trees. Similarly, *coverage* was measured as the percentage of simulations in which the 95% HPD of the estimated time interval length $(t1 - t2)$ contained the true (simulated) time of the single-pulse MEE, *t.*

Comparison with the single-pulse CoMET model

Simulation study: Because each analysis was time-consuming, we made a random selection of ten extant trees for each of the 10 parameter combinations, totaling 240 simulated trees. We conditioned the CoMET model on taxa survival, and used the function *tess.analysis* from TESS (Höhna et al. 2015) to estimate the number and magnitude of rate shifts in μ and λ and the number of MEEs from the phylogeny (*estimateNumberMassExtinctions*, *estimateMassExtinctionTimes* = TRUE). We used default parameter settings for the priors of the three independent CPP processes: *numExpectedMassExtinctions* and *numExpectedRatechanges* for λ and μ were set to equal to two; this assigns a 50% probability to zero MEEs and zero rate changes (Höhna et al. 2015). A *Beta*(α=5, β=95) distribution was used as the prior for the mass extinction survival probability, which assigns an expected value of $\rho = 0.1$ (a MEE with 0.9 intensity). The prior distributions for the speciation and extinction rates were estimated using an empirical Bayesian approach with the function *empiricalHyperPriors*=TRUE. In this approach, a short preliminary analysis is performed in CoMET under a constrained, constant birth-death model to estimate reasonable values for the hyperprior distributions of λ and μ, which are then used in a longer unconstrained CoMET analysis to estimate the marginal posterior distributions of all parameters (May *et al.* 2016). As with BDSKY above, we assumed complete taxon sampling at present ($\rho = 1$). For each tree, the model was run until the ESS for every parameter reached a value of 500. Bayes Factor comparisons were employed to evaluate the marginal likelihoods of competing models for the timing of the MEEs, with significance values following in Kass and Raftery (1995):

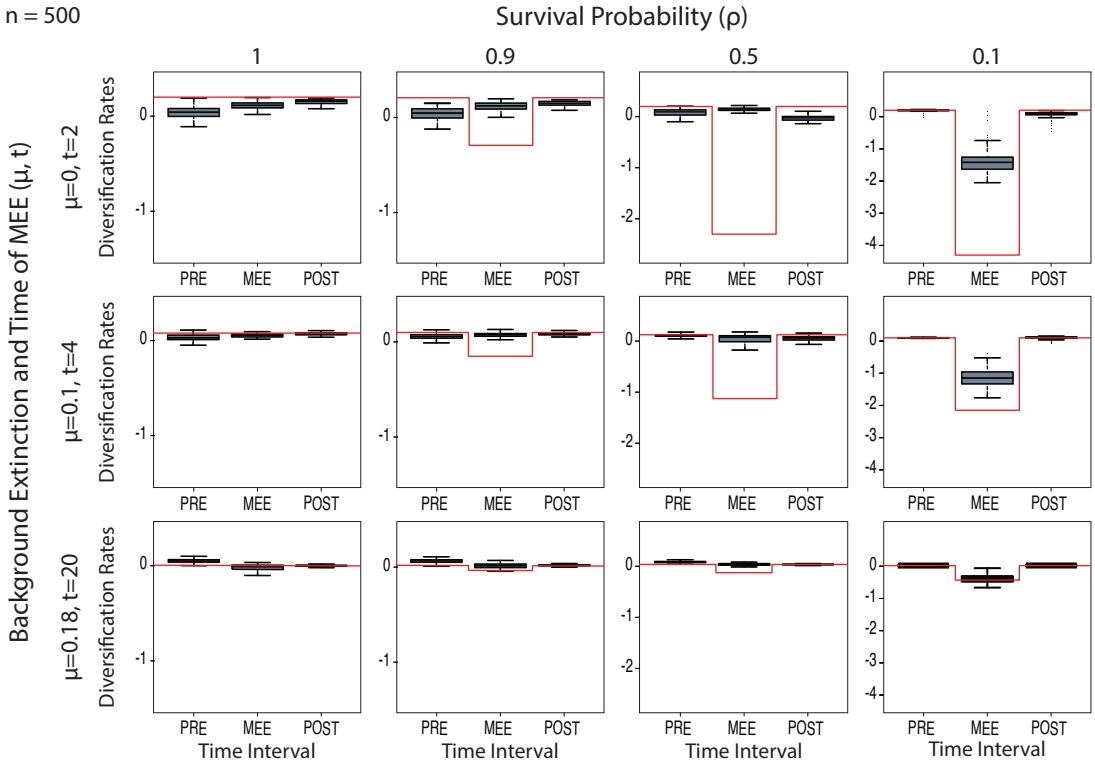


Figure 2: Detection of MEEs under the BDSKY Model C through sequential changes in the magnitude of the diversification rate (diversification = λ - μ) under varying levels of μ and mass extinction survival probability, ρ. The red line represents the true (simulated) value; this has been adjusted for ρ = 0.1 in the MEE time interval to reflect the effect of the MEE (see text). The boxplots show the variance in the estimated value across simulated trees, depicting the mean of the means of all trees (thicker dark line), the 75-24% interquartile ranges (shaded box) and the post extreme data points (whiskers). Notice that in the high-extinction scenario (ρ = 0.1), the diversification rate becomes negative in the second time interval, followed by recovery to positive values in the next interval, signaling the presence of the MEE (μ > λ), but that this change is not observed in the control scenario (with no mass extinction, ρ = 1). Figure S3 and Figure S4 show the same results for Models A and B, respectively.

BF > 2 (positive support), and BF > 6 (strong support).

Empirical study: To compare the robustness of BDSKY and CoMET against a real empirical dataset, we analyzed the conifer time tree of Leslie *et al*. (2012: 342 taxa, 78% taxon sampling), which was also used in May *et al.*'s (2016) study. For BDSKY, we ran models B and C setting the number of rate shifts to vary between two and five to emulate CoMET in allowing for multiple sequential MEEs. We also cut off ("masked") part of the tree length by instructing the model to start searching for a MEE after ~6%, ~11%, and ~18% of the tree root age (340 Mya): 320, 300 and 280 Mya; this was done in order to remove parts of the tree with very little information regarding MEEs. The initial rate shift values were set to be equally-spaced across the tree length. Prior distributions for μ and λ rates followed May *et al*. (2016, cf. Supplementary Information S14): a *lognormal* distribution with a standard deviation of 0.02 and mean=0.09 for μ and 0.16 for λ. Analyses were run until the ESS value reached 200 for each

parameter; analyses that did not converge after 1000 million iterations were discarded. All data results from this study are deposited in the public repository dryad, https://doi.org/10.5061/dryad.qv10c62.

RESULTS

Impact of parameter settings in simulated phylogenies

Figures S1 and S2 show the simulated trees –full (including extinct and extant taxa) and reconstructed (extant taxa only)– under the different parameter-combination settings. Comparison of the lineage-through-time (LTT) plots among scenarios reveals several aspects that we expect from analytical considerations (e.g. Gernhard 2008). The background extinction (ratio of a varying μ over constant λ) has a large influence over: the root age: as μ increases, the root age in the full and reconstructed trees increases; the root age variation: as μ increases, the variation increases, and the variation of the root age between the full and reconstructed trees: as μ increases, this difference also increases.

Parameter estimation under the BDSKY model

We compare below the performance of the three BDSKY analyses (Models A to C) in terms of how well they answer the questions posited in the Introduction:

(i) Can we detect the phylogenetic signal of a MEE in the extant phylogenetic tree through changes in the diversification rate? Results from the control scenario (with no mass extinction, $\rho = 1$) showed that the percentage of false positives or Type I error was very low in the three models: HPDn values were 0 for all models (Model C: Figure 2, Model A: Figure S3, Model B: Figure S4). Precision and accuracy for estimates of the diversification rate were best for Model C, followed by Model A (Table S2; Figure 2, Figure S3); both models were also able to capture the true simulated value within the 95% HPD interval in 100% trees (full coverage, Table S2). The time-constrained Model B showed the lowest values for accuracy, precision, and coverage, especially in the first (pre-MEE) and second (MEE) intervals (Figure S4, Table S2). The MEE interval is most likely too short, containing a very low number of nodes, and thus the method cannot detect any significant result for that time interval. The percentage of false negatives or Type II error in trees simulated under varying levels of MEE survival probability (ρ) was highest for Model B, which showed no significant decrease in diversification rate in the MEE time interval (i.e. HPDn was always 0, Table S2), irrespective of the value of extinction μ (Figure S4). Accuracy, precision and coverage were also the lowest among the three models when $\rho < 1$ (Table S2). Because of this failure to detect the MEE under varying settings, Model B was

discarded and will not be further commented upon. Between the other two models and for the high-intensity mass extinction scenario (ρ = 0.1) and N = 500, Model C (constrained speciation) showed the highest power in detecting the MEE irrespective of the value of μ, showing negative values for the diversification rate in the MEE time interval, with a rebound to positive values in the third (Figure 2), an indication of the ability of the model to detect the MEE. Even when the coverage was 0 (the true value was not contained within the 95% HPD interval), as in the case of μ = 0, the model was returning a negative value for the diversification rate in the MEE time interval in 84% of the trees that did converge (HPDn = 0.84, Table S2), and this percentage became higher with increasing values of μ (93%, 97%). Accuracy, precision and coverage also increased with higher values of μ, and were the highest for μ = 0.18 (Figure 2, Table 1).

As expected, the percentage of false negatives (i.e. failing to detect the MEE) increased as the mass extinction intensity ($1 - \rho$) decreases (Figure 2), with HPDn values generally close to zero for ρ = 0.5 and ρ = 0.9 (Table 1). Accuracy, precision, and coverage were lower for ρ = 0.5 than for ρ = 0.1 (except for the MEE interval with μ = 0), whereas ρ = 0.9 showed values very similar to the null model (Table 1). Similarly, a lower number of extant taxa in the reconstructed phylogeny (N) translates into a decrease in the power to detect the MEE, especially for μ = 0 (Figure 3).

The full parameterized Model A (Figure S3) performed worse than Model C, showing HPDn values close to zero (Table S2). A trend towards negative values can be observed for the scenario with the lowest MEE survival probability (ρ = 0.1) and N = 500, signaling the presence of a MEE (Figure S3). But precision and coverage were considerably worse than in Model C (broader 95% HPD intervals, Table S2), and accuracy was also lower (Table S2). The ability of the model to detect the MEE was similarly low for higher values of ρ (0.9, 0.5, Figure S3) and smaller values of N (100, 200, Figure S5).

(ii) Can we accurately estimate the time interval of the MEE from an extant phylogenetic tree? Figure 4 shows that Model C performed well in estimating the time of the rate shifts bounding the MEE interval within which the MEE was expected to occur, for the high-intensity scenario (ρ = 0.1) and varying values of μ. Coverage was always equal to 1, and 95% HPD intervals (precision) were narrow, especially for μ = 0 and 0.1 (Table 1). However, this power notably decreases with higher values of MEE survival probability: the variance in the two rate shift estimates pre- and post-MEE overlaps across age values for ρ = 0.9 and ρ = 0.5 (Figure 4), indicating the failure of the model to capture the rate shift times under low-intensity MEE scenarios. Likewise, for N < 500, the model is unable to estimate the

Table 1. Summary statistics for the parameters in the Birth–Death Skyline Model C (N = 500; see text for more detailed description).

		ρ = 1						ρ = 0.9						ρ = 0.5						ρ = 0.1					
		λ			μ			λ			μ			λ			μ			λ			μ		
	Time interval	Acc	Prec	Cov	Acc	Prec	Cov	Acc	Prec	Cov	Acc	Prec	Cov	Acc	Prec	Cov	Acc	Prec	Cov	Acc	Prec	Cov	Acc	Prec	Cov
μ = 0.0	Pre	−0.02	0.07	0.91	−0.03	0.59	0	−0.02	0.07	0.95	−0.03	0.58	0	−0.03	0.1	0.8	−0.08	0.49	0	−0.03	0.08	0.77	−0.19	0.1	0
	MEE	–	–	–	−0.11	0.4	0	–	–	–	−0.11	0.38	0	–	–	–	−0.14	0.37	0	–	–	–	5.84	1.72	0
	Post	–	–	–	−0.14	0.27	0	–	–	–	−0.13	0.26	0	–	–	–	0.03	0.4	0	–	–	–	−0.05	0.39	0
μ = 0.1	Pre	0	0.07	0.96	0.05	0.46	1	0	0.08	0.99	0.05	0.44	1	−0.01	0.09	0.96	0.02	0.31	1	0	0.08	0.97	0	0.09	0.97
	MEE	–	–	–	0.03	0.33	1	–	–	–	0.03	0.33	1	–	–	–	0.09	0.49	1	–	–	–	3.49	1.41	0.24
	Post	–	–	–	0.01	0.19	0.99	–	–	–	0.02	0.21	1	–	–	–	0.07	0.27	0.84	–	–	–	0	0.22	0.99
μ = 0.18	Pre	0	0.07	0.97	0.12	0.32	0.99	0	0.07	0.95	0.11	0.32	0.98	0	0.07	0.97	0.12	0.28	0.99	0	0.07	0.94	0.16	0.08	0.93
	MEE	–	–	.	0.18	0.42	1	–	–	–	0.17	0.4	1	–	–	–	0.17	0.35	1	–	–	–	1.02	0.77	0.94
	Post	–	–	–	0.16	0.14	0.99	–	–	–	0.16	0.14	0.99	–	–	–	0.17	0.15	0.97	–	–	–	0.16	0.12	0.94

Abbreviations: "μ," value of extinction rate in simulations";," survival probability in simulations; mass extinction intensity = (1 −ρ). "Acc," the mean of the means of the estimated parameters across trees; "Prec," the mean of the width of the 95% high posterior density (HPD) credibility interval across trees; "Cov," coverage, percentage of simulated trees where the 95% HPD credibility interval contained the true parameter value.

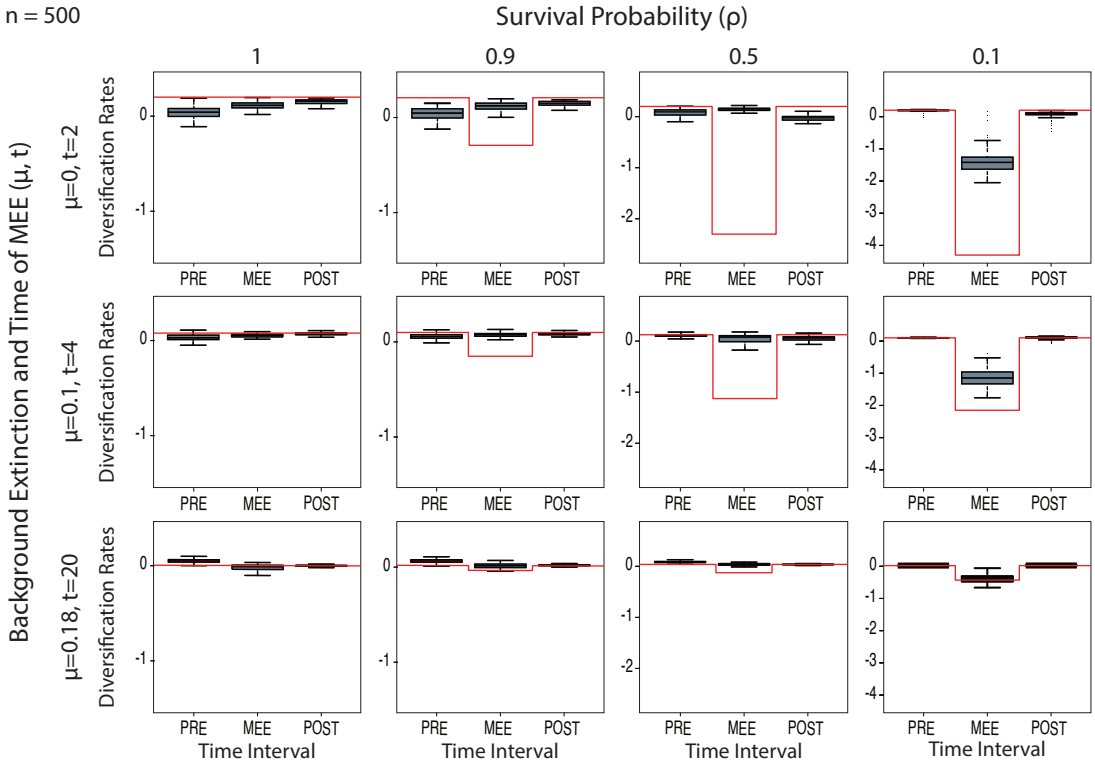


Figure 3: Detection of MEEs through interoperating changes in the magnitude of the diversification rate in Model C under varying levels of N (number of taxa). All other conventions follow Figure 2.

MEE time interval with any accuracy or precision (Figure S6). For the null model (ρ = 1), the time length covered by the variance in these estimates (boxplots in Figure 4) spans the entire range of root ages across simulated trees, as can be expected in the absence of a MEE. The full Model A (Figure S7, S8) again performed worse than Model C in terms of precision and coverage (Table S2): it failed to estimate the rate shift times bounding the MEE for any parameter combination of μ and ρ (Figure S7), as well as for the low values of N (100, 200, Figure S8).

(iii) Can we provide reasonable estimations for λ and μ pre- and post-MEE?

Since Model C was the only model that could successfully detect and estimate the time interval of the MEE (for N = 500 and ρ = 0.1), we focus here on this model to evaluate the performance in parameter estimation of μ and λ.

Figure 5 shows the variance in the estimates of λ and changes in the magnitude of μ across the three time intervals, for different simulation scenarios. Estimates of λ were good across varying settings of μ and ρ, especially regarding accuracy and precision (Table 1); coverage was lower for low values of extinction (Table 1). Estimates of λ were also good for the control scenario in terms of accuracy, precision, and coverage (Table 1).

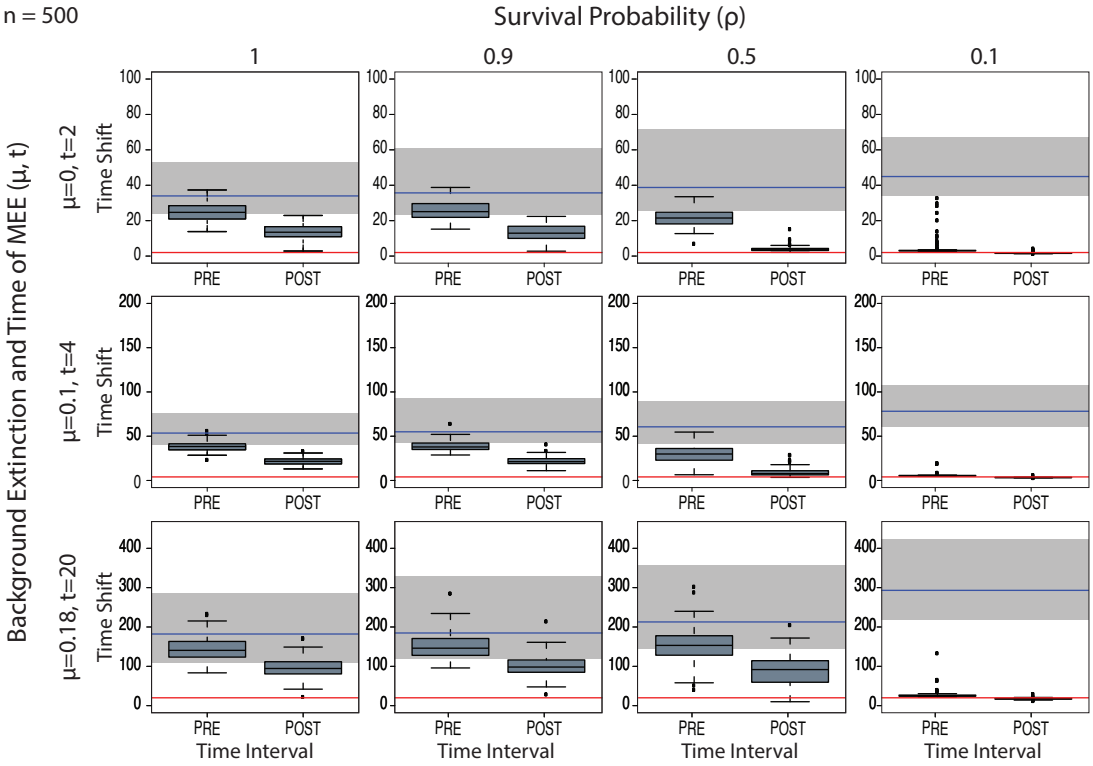


Figure 4: Estimates of the pre-MEE and post-MEE rate shift times bounding the second time interval, for Model C and under varying levels of μ and mass extinction survival probability, ρ. The red line indicates the value of the true (simulated) time of the MEE, *t*. The time is shown from present to past: left boxplot (post-MEE) corresponds to the rate shift between time intervals "MEE" and "post-MEE", after which diversity is expected to recover; right boxplot (pre-MEE) corresponds to the rate shift between time intervals "pre-MEE" and "MEE", after which the MEE is expected to have occurred. The grey bar indicates the variance in root ages across tree simulations, while the blue line shows the mean of this range. Notice the low variance and the small difference between the two boxplots (pre- and post-MEE rate shifts), indicating that the time-slice model is able to locate the MEE even when modelled as a nearly single-pulse (instantaneous) event. See Figure S6 for the results with varying values of N (100, 200, 500). Figure S7 and Figure S8 show the equivalent results of this analysis for Model A.

For estimates of μ, in general the best results were obtained with the low mass extinction survival probability scenario (ρ = 0.1), especially for the pre- and post-MEE time intervals, and for scenarios with moderate (μ = 0.1) or high (μ = 0.18) background extinction. For μ = 0, the coverage (95% HPD containing the true value) was very low (Table 1; true value is red line in Figure 5, ρ = 0.1), and precision and accuracy were generally worse. However, the model was able to recover the signal of an increase in background extinction in the MEE time interval for the majority of trees (see Figure 5, ρ = 0.1).

As the mass extinction survival probability, ρ increases, the power to accurately measure changes in μ also decreases; no significant differences in variance were observed with respect to the null model (ρ = 1), with the rate of μ over- or underestimated across time intervals (Figure 5, ρ ≠ 0.1). Lowering the number of extant taxa also

resulted in a lower ability to detect changes in magnitude in μ (Figure S9). Yet, for N = 200 and moderate/high background extinction, Model C was still able to capture the increase in the extinction rate for the MEE time interval, and estimates of λ and of μ for the post- and pre-MEE time intervals were relatively accurate (Figure S9).

Comparison against single-pulse CoMET

Simulation Study: Table 2 summarizes the differences and similarities between the BDSKY and CoMET models in terms of performance with the set of simulated trees (see Figure S10-S15 for CoMET). The two models have a similar frequency of Type II error as defined here, failing to detect the MEE through changes in the diversification rate (HPDn), and they exhibit no Type I error, i.e. false detection of a MEE where none was simulated, ρ =1 (Table 2). The BDSKY model performed slightly better in detecting the MEE under the high-intensity mass extinction scenario (ρ = 0.1) and N=500, whereas CoMET did the same for the moderate-intensity scenario (ρ = 0.5); CoMET also performed better for smaller phylogenies with N=100 and N=200. The same pattern can be observed in the location of the MEE, i.e. estimating the timing of the MEE: both models succeeded with large phylogenies (N=500) under the high-intensity scenario (ρ = 0.1), but CoMET performed better under for the moderate-intensity scenario (ρ = 0.5); CoMET also performed better for smaller phylogenies with N=100 and N=200. The same pattern can be observed in the location of the MEE, i.e. estimating the timing of the MEE: both models succeeded with large phylogenies (N=500) under the high-intensity scenario (ρ = 0.1), but CoMET performed better under moderate/low-intensity mass extinction scenarios (ρ = 0.5, 0.9) and with smaller phylogenies (N=100, 200). There is one interesting difference, however. Whereas in BDSKY, the timing of the MEE was identified through the rate shift times of μ –in accordance with the time-slice model– in CoMET, MEEs were often located through the timing of rate shifts for λ (i.e., as a significant decrease in speciation rates) rather than through the specific parameter used in CoMET to detect MEEs, mass extinction times. If the mass extinction times parameter succeeded to locate the MEE, the Bayes Factor comparisons were often not significant (2 < BF < 6, Table 2). In general, BDSKY showed better accuracy than the CoMET model in the estimation of μ and λ, irrespective of the MEE intensity (1-ρ) and the size of the phylogeny (N). CoMET performed worse in scenarios with no extinction (μ = 0), systematically overestimating μ or over/underestimating λ. Both BDSKY and CoMET showed the best behavior under scenarios with moderate extinction (μ = 0.1, Table 2).

Empirical study: Figure 6 shows the results of the BDSKY and CoMET models with the empirical conifer dataset of Leslie et al.

CHAPTER 1

Table 2. Table summarizing the performance of the time-slice BDSKY Model C, and the single-pulse CoMET model (May et al. 2016) for the simulated set of phylogenies that converged in BEAST2 (see Table S1). Color code: "White" indicates the success of the model in estimating the parameter value or detecting the MEE event; "green" indicates the failure of the model; "yellow" indicates mixed results. See footnotes for an explanation (the corresponding Figure numbers illustrating these results are given under each header; posterior probability estimates and accuracy values for each parameter are given in Table S3).

	λ estimation		μ estimation		MEE detection		MEE time estimation	
Model Settings (N, ρ, μ, t)	BDSKY (5, S9)	CoMET (S14, S15)	BDSKY (5, S9)	CoMET (S14, S15)	BDSKY (2, 3)	CoMET (S10, S11)	BDSKY (4, S6)	CoMET (S11, S13)
(500, 0.1, 0, 2)	a	a^1	a^2	a^1	b^*	b^1	c	c – λ
(500, 0.1, 0.1, 4)	a	a^{1*}	a	a^{1*}	b	b	c	c – λ
(500, 0.1, 0.18, 20)	a	a	a	a	b	b	c	c – λ c^2 – μ c^2 – MEE
(500, 0.5, 0, 2)	a	a	a^1	a^1	b^2	b^2	c^2	c^1 – λ
(500, 0.5, 0.1, 4)	a	a^*	a	a^*	b^2	b^1	c^2	c – λ
(500, 0.5, 0.18, 20)	a	a	a	a	b^2	b^2	c^2	c^2
(500, 0.9, 0, 2)	a	a	a^1	a^1	b^2	b^2	c^2	c^2
(500, 0.9, 0.1, 4)	a	a	a	a^*	b^2	b^2	c^2	c^2
(500, 0.9, 0.18, 20)	a	a	a	a^*	b^2	b^2	c^2	c^2
(500, 1, 0, 2)	a	a	a^2	a^1	d	d	d	d
(500, 1, 0.1, 4)	a	a	a	a^*	d	d	d	d
(500, 1, 0.18, 20)	a	a	a	a	d	d	d	d
(100, 0.1, 0, 2)	a^{1*}	a^2	a^1	a^1	b^2	b^1	c^2	c – λ
(100, 0.1, 0.1, 4)	a^*	a^*	a^2	a^*	b^2	b^2	c^2	c^1-λ c^2– MEE
(100, 0.1, 0.18, 20)	a	a	a^{**}	a^*	b^2	b^2	c^2	c^2– MEE
(200, 0.1, 0, 2)	a^1	a^{1*}	a^1	a^1	b^2	b^{1*}	c^2	c – λ
(200, 0.1, 0.1, 4)	a	a^1	a^{**}	a^{1*}	b^2	b	c^2	c – λ
(200, 0.1, 0.18, 20)	a	a	a^{**}	a	b^{1*}	b^2	c^2	c– MEE
(100, 1, 0, 2)	a	a^1	a^1	a^1	d	d	d	d
(100, 1, 0.1, 4)	a	a^*	a^*	a^*	d	d	d	d
(100, 1, 0.18, 20)	a	a^*	a^*	a^*	d	d	d	d
(200, 1, 0, 2)	a	a^1	a^2	a^2	d	d	d	d
(200, 1, 0.1, 4)	a	a	a^*	a^*	d	d	d	d
(200, 1, 0.18, 20)	a	a^*	a	a^*	d	d	d	d

a: Simulated (true) value falls within 95% HPD (BDSKY) or Credible Interval (CoMET). (*)Large 95% HPD interval width (≥ 0.05 between lower and upper boundary). (**) Large only for post – MEE interval.

a^1: Under/Overestimation of true value (falls outside the 95% HPD or Credible Interval). (*) Mean overestimated but the true value falls within 95% HPD (BDSKY); under/overestimation only observed in part of the tree length (CoMET).

a^2: True value falls within 95% HPD (BDSKY) or Credible Interval (CoMET), but only in either the pre- or post-MEE interval.

b: Success in detecting MEE: Mean and 95% HPD (BDSKY)/Credible Interval (CoMET) of the diversification rate estimate fall below 0 ($r < 0$) at MEEE ("MEE interval" in BDSKY) and goes back to simulated values after MEE ("post-MEE interval"). (*) Only 84% of HPDn < 0 for the percentage of simulated phylogenies that converged.

b^1: Weak detection of MEE: Diversification rate decreases in pre-MEE interval (BDSKY) or at MEE (CoMET), but mean and/or 95% HPD/Credible Interval of the diversification rate is not negative. (*) Only part of the HPD falls below 0.

b^2: Type II error: Failure to detect the MEE through the diversification rate.

c: Good estimation. MEE time correctly bounded by rate shift times in μ (BDSKY). MEE time correctly identified by significant Bayes Factor comparisons (BF > 6) of λ shift times (c – λ) or single-pulse MEE times (c-MEE) (CoMET).

c^1: Weak estimation (CoMET): MEE time correctly identified by non-significant BF tests (2 < BF < 6) of rate shift times in λ (c^1-λ), or in (c^1-μ), or single-pulse MEE times (c^1-MEE).

c^2: Failed estimation. MEE time incorrectly identified by rate shift times in μ (BDSKY) or by non-significant BF tests of μ shift times (c^2-μ) or single-pulse MEE times (c^2-MEE) (CoMET).

d: No Type I error. No MEE is detected in the control scenario ($\rho = 1$).

Abbreviations: λ estimation, power to estimate speciation rate; μ estimation, power to estimate extinction rate; MEE detection, power to detect the MEE through interoperating changes in the diversification rate; MEE time estimation, timing of MEE detected through successive shifts in extinction rate estimates (BDSKY) or through shifts in speciation rate (CoMET).

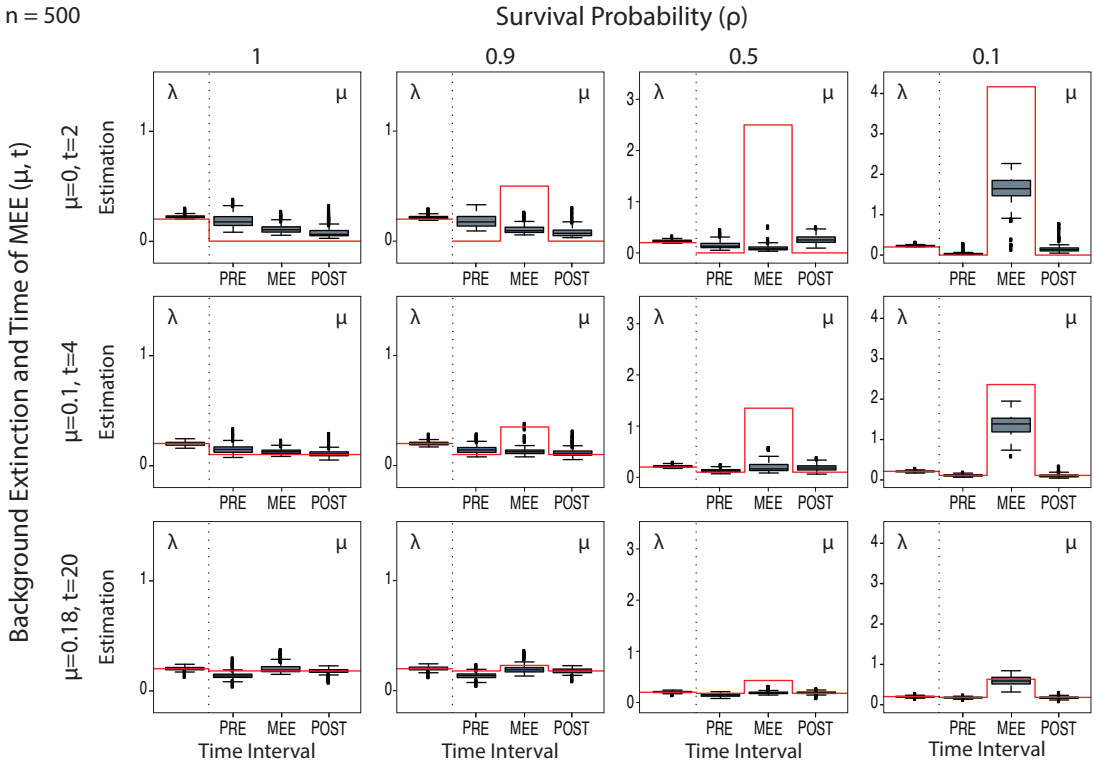


Figure 5: Estimation of changes in magnitude of μ across time intervals for Model C under different values of background extinction and MEE survival probability, ρ. In this model, λ is estimated but assumed constant over time; i.e. MEEs are only detected through changes in the magnitude of μ. The red line represents the true simulated value; this has been adjusted for ρ = 0.1 in the MEE time interval to reflect the effect of the MEE (see text). Notice the large increase of μ (> 1) in the MEE time interval for ρ = 0.1, indicating the presence of the MEE, while this effect is not seen in the control scenario (ρ = 1). See Figure S9 for the results with varying values of N (100, 200, 500).

(2012). CoMET detected a significant MEE (BF > 10) at c. 23 Mya, as in May et al. (2016). BDSKY located a drop in net diversification to $r < 0$ at 43 (78-23) Ma and an increase in net diversification to $r > 0$ at 3 (10-0) Ma, given we allow for two rate shifts. If we allow for five rate shifts, the drop happens within roughly the same HPD interval (70-20 Ma) with the median being at 30 Ma (instead of 43 Ma). Thus, the estimated period of negative diversification spans several decades of millions of years. The instantaneous MEE of May et al. is estimated to be at 23 Ma, and thus falls within the period of negative net diversification that we estimate. In contrast, the other, "potential" mass extinction events reported by May et al. 2016 at 173 Mya and 77 Mya ($2 < BF < 6$) are not recovered by BDSKY as true MEEs according to our definition ($r < 0$, $\varepsilon > 1$), but as low-magnitude rate-shifts in μ.

DISCUSSION

Time-slice versus single-pulse approach to detect mass extinction events

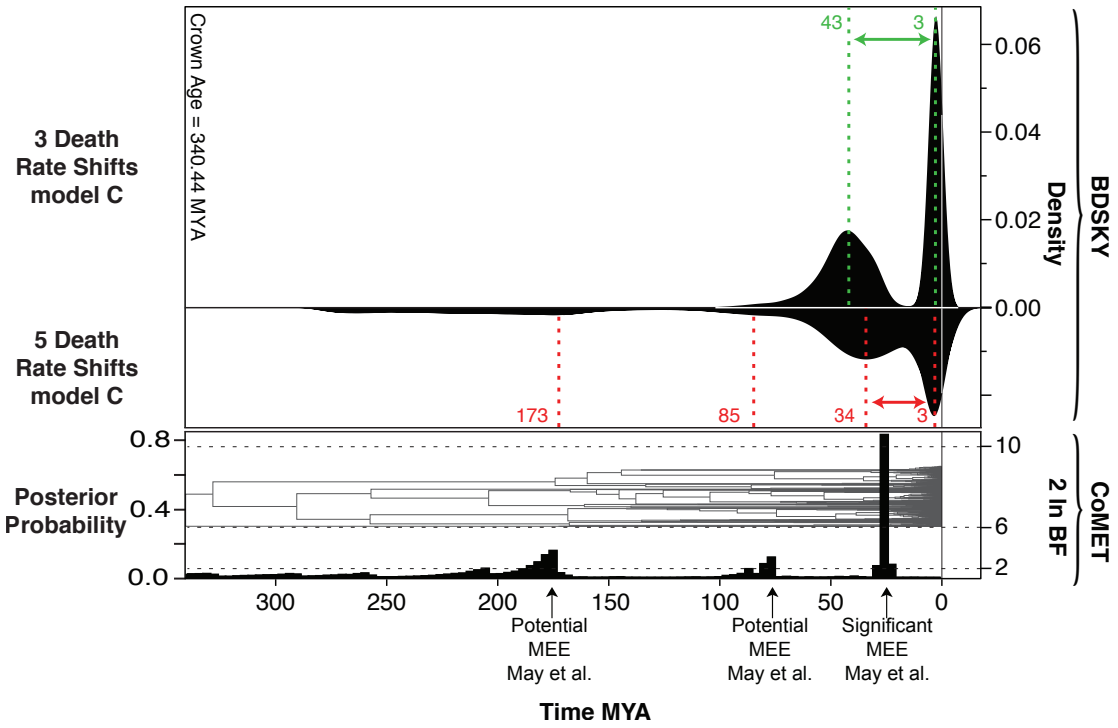


Figure 6: Comparison of the BDSKY and CoMET model performance with the empirical conifer dataset (Leslie *et al.* 2012): BDSKY was run under Model C. We explored models with one and multiple rate shifts (shown here are the best models with two and five rate shifts). CoMET was run under the same settings used by May *et al.* (2016). CoMET detected a major event of mass extinction at c. 23 Mya. BDSKY indicated a drop in the net diversification to $r < 0$ at c. 43 or 34 Ma, dependent on the number of rate shifts, and an increase in net diversification to $r > 0$ at 3 Ma. BDSKY, suggests that the two non-significant MEEs ($2 < BF < 6$) detected by CoMET at 173 and 77 Mya (May *et al.* 2016), are not true MEEs ($\varepsilon > 1$), but rather low-magnitude rate-shifts in μ.

Evolutionary episodes of hyperdiversification –i.e. defined as speciation rates within a lineage that are significantly higher than those expected under the background diversification rate of their encompassing clade– have long attracted the attention of biologists (Hughes and Eastwood 2006; Valente *et al.* 2010) and analytical systematists (Rabosky 2006; Alfaro *et al.* 2009; Rabosky 2014) because of their potential links to "key innovations" (morphological novelties) or the colonization of novel environments leading to increased species fitness, "key opportunities" (Wiens *et al.* 2010; Donoghue and Sanderson 2015).

By contrast, episodes of high extinction rates such as mass extinction events (MEE), have traditionally received less attention in the phylogenetic literature because of the difficulty of measuring a process that removes rather than generates diversity (Pyron and Burbrink 2012; Sanmartín and Meseguer 2016). However, these MEEs form a key element of the paleontological record as responsible for major ecosystem reordering and change (Raup 1979; Sepkosky 1982; Purvis 2008; Benton 2009),

and recently have regained importance in the context of human-induced biodiversity loss (Barnosky *et al.* 2011a, b).

Unlike speciation or background extinction rates, which are assumed to depend on species biotic traits or a clade's ecology (Purvis 2008; Ezard *et al.* 2011), MEEs are often linked to abiotic factors, i.e. long-term environmental changes or catastrophic, geological events whose effects are felt across many lineages (Pyron and Burbrink 2012; Sanmartín and Meseguer 2016). Because of this, MEEs are often modeled in the phylogenetic literature as tree-wide events that act simultaneously across clades in contrast to events that are clade-specific (Stadler 2011a, b). Indeed, most macroevolutionary approaches model MEEs as random instantaneous extinction events, in which the standing diversity is reduced by a fraction equal to the magnitude or intensity of the mass extinction, 1–ρ, (Harvey *et al.* 1994; Stadler 2011b; May *et al.* 2016).

Modeling MEEs under this single-pulse approach (i.e. through the parameter ρ) –as in the birth-death models implemented in *TreePar* and TESS– has the advantage that one can estimate the magnitude or intensity of MEEs, and it allows for the statistical testing of time-specific MEE hypotheses (Stadler 2011a; May *et al.* 2016; Sanmartín and Meseguer 2016). Yet, the single-pulse approach stands in contrast with the paleontological literature, where MEEs are defined in terms of intensity and duration, where unusually high background extinction rates take place over a (geologically) short time period, followed by a recovery or a return to positive net diversification rates (Raup 1979). We showed here that the Bayesian birth-death skyline (BDSKY) model (Stadler *et al.* 2013) – developed initially to trace the spread of viral or bacteria infections over time – can be used to detect MEEs under the "time-slice" approach: MEEs are identified and located through a combination of negative diversification rates and two sequential rate shifts in the extinction rate occurring in a significantly short period of time (Figure 2; Table 1).

A second advantage of the time-slice approach relates to the issue of parameter non-identifiability. Joint estimation of instantaneous tree-wide changes in dispersal and extinction rates (i.e., rate shifts in λ and μ) and MEE single-pulse events (ρ) is not possible because the three parameters are modeled with the same likelihood function, and in fact the model becomes non-unidentifiable. This applies to *TreePar* when allowing for both single-pulse and time-slice MEEs, i.e. through diversification rate changes (this option is not recommended in the R package documentation due to parameter correlations). To escape from this paradox, CoMET, which can also be used to model both types of MEEs, uses a hierarchical Bayesian approach in which MEEs are estimated by marginalizing over other nuisance parameters such as all possible

instantaneous shifts in speciation and extinction rates (May *et al*. 2016). Therefore, CoMET may be seen as a more general and powerful tool than BDSKY. While CoMET solely returns single-pulse MEEs, it additionally also outputs changes in speciation and extinction rates, which may be regarded as equivalent to returning time-slice MEEs. However, the risk of diluting the signal of the MEE by explaining it partially as single-pulse and partially as a time-slice exists. This can be observed in our results for the simulation data in CoMET, where we set up the MEE model to be a single-pulse (i.e. recover the MEE through the parameter *mass extinction times*); however, we mainly recovered MEEs under the time-slice approach, i.e. as *speciation rate changes* (Figure S12-S13, Table 2).

In BDSKY, we investigate the posterior distribution of the extinction rate through time and thus detect the presence of MEEs in the phylogeny as time periods with elevated extinction rates. There is no need to mathematically disentangle MEEs from background extinction and speciation rate shifts (Stadler 2011c; May *et al.* 2016) because we estimate rate shifts in background extinction and MEEs within the same continuous-time framework (Figure 1b). This can be directly seen with the example of the conifer dataset (Figure 6): BDSKY identifies the increase in extinction rates detected by CoMET at c. 23 Mya as a genuine MEE ($\varepsilon > 1$), whereas the other low-magnitude events (173 and 77 Mya in CoMET) are identified as periods of high background extinction rates with $\varepsilon < 1$. These changes in extinction rate were picked up only when the BDSKY model was allowed to explore multiple rate shifts, indicating that as the number of rate shifts increases, BDSKY becomes more sensitive to the influence of low-magnitude changes in μ. Yet, parameter identifiability issues can also affect the BDSKY model. Our simulations were generated under the single-pulse approach to MEEs, as instantaneous events, and then estimated under the MEE time-slice approach in BDSKY; notice the low variance and the small difference between the pre-MEE and post-MEE values of rate shifts in Figure 4. Since pulses can be seen as the limit of time-slice models, then, at least for extremely short time slices with very high extinction (which would be well-approximated by a pulse model), BDSKY could get into the same problems of parameter non-identifiability as CoMET or *TreePar*. In future studies, it would be interesting to test the performance of the BDSKY model for longer-duration MEEs scenarios.

It is important to emphasize that all three methods, CoMET, *TreePar* and BDSKY, rely on the same likelihood function for estimating speciation, extinction and MEEs, and therefore conceptually allow for both types of MEEs, namely single-pulse and time-slice scenarios. However, they differ in their statistical framework and the type of prior

distributions used for detecting MEEs, which conditions their performance. CoMET (Bayesian) and *TreePar* (ML) were specifically designed to identify single-pulse MEEs (also termed *explicit* MEEs in May *et al.* 2016); the time-slice MEEs would not be reported as MEEs but as "regular" changes in extinction and speciation rates. Instead, BDSKY was developed to detect rate changes (Stadler *et al*. 2013) and is thus better tuned to detect MEEs under the time-slice approach. In particular, CoMET focuses on the detection of instantaneous events, while BDSKY focuses on the time interval during which MEEs occur.

Comparing the performance of BDSKY and CoMET

Our simulation study reveals that BDSKY and CoMET performed well and suffered from similar low rates of Type I (false discovery rate) and Type II error (failing to detect the MEE) for the control ($\rho = 1$) and high-intensity MEE ($\rho = 0.1$) scenarios. For the small (N=100) and medium-sized (N = 200) phylogenies, CoMET was more robust to Type I and Type II errors in the detection and location (timing) of the MEE than BDSKY. On the other hand, BDSKY performed better than CoMET in the estimation of the rate of speciation (λ) and of changes in the magnitude of extinction (μ) across time intervals, in particular for N = 500 and $\rho = 0.1$.

Overall, CoMET has a larger number of parameters than BDSKY because it models the frequency of MEEs and speciation and extinction rate shifts through three independent CPP processes, and also includes the survival probability parameter ρ (May *et al.* 2016). CPP models are known to be sensitive to the choice of priors and suffer from parameter identifiability issues (Rannala 2002). Because of this, May *et al*. (2016) advocated the use of strongly informative (biologically grounded) priors on some parameters such as the frequency or the intensity of MEEs to increase the power of CoMET in detecting these events. For example, one can use the paleontological record (Benton 2009) to inform the MEE intensity ($1-\rho$), which should be c. 90%, or the expected number of MEEs (λ_M) in the CPP prior, as in the conifer example (Höhna *et al*. 2015; May *et al.* 2016). A prior expectation on the number of rate shifts in speciation and extinction (λ_B and $\lambda_{D)}$ is more difficult to justify on biological grounds, and often given the same value as the one on the expected number of MEEs (*numberExpectedratechanges* = *numberExpectedMassExtinctions* = 2); Höhna *et al*. 2015). For the speciation and extinction rates, an empirical Bayesian approach is used to parameterize the hyperpriors. May *et al*. (2016) argued that the use of Bayes factor comparisons cancels out these prior assumptions, which are only used to speed up the convergence of the MCMC search but should not have an effect on the eventual

conclusions (Höhna *et al*. 2015). Yet, due to the issues of parameter non-identifiability (above), the results from CoMET may still be dependent on the choice of hyperprior being used in the analysis. Moreover, because MEEs in CoMET are estimated by integrating over speciation and extinction rate shifts, which are considered as nuisance parameters in the model, May *et al.* (2015) originally warned that researchers should be cautious not to over-rely on the values of these parameters (tree-wide diversification rate shifts), for which we have also less biological information compared to MEEs. This could explain the poor performance of CoMET in estimating the magnitude of rate changes in λ and μ (Table 1). In contrast, BDSKY is unaffected by this issue because MEEs are estimated through changes in the background extinction rate, and therefore MEEs and temporal changes in extinction rates are treated in the same way (Condamine *et al*. 2013; Figure 4).

It is important to notice, however, that we are referring here to the BDSKY Model C, in which we constrained the speciation rate (λ) to be constant across time intervals; Model A, allowing speciation and extinction rates to vary, as CoMET does, performed significantly worse. The assumption of constant speciation rates might seem unrealistic and hard to justify on biological grounds. Yet, in an often-cited study, Ezard *et al.* (2011) found that speciation rates in the fossil record were mostly shaped by biotic factors such as diversity dependence, whereas fossil-based extinction rates responded mainly to abrupt abiotic perturbations, such as major geological or climatic changes. Therefore, it might be appropriate to detect time-slice MEEs through the extinction rate parameter, which is (potentially) more sensitive to these environmental changes than the speciation rate.

In all, both BDSKY and CoMET, have their strengths and drawbacks and could be use side-by-side. CoMET performs best as a method for testing time-specific hypotheses on MEEs, when we have some prior information on their presence, while BDSKY can be seen as an exploratory model to search for the signal of potential MEEs in a phylogeny in the absence of such data. This can be exemplified again in the conifer phylogeny. CoMET detected a major episode of mass extinction at c. 23 Mya, as well as the weak signal of other increases in extinction rate in the distant past. Though May *et al*. (2016) regarded this MEE as a possible artifact of biased divergence time estimates, it is temporally congruent with the Late Oligocene Warming Event (LOWE, 26.7-23.5 Mya, Zachos *et al*. 2001), a widespread major warming pulse that accompanied the closing of the eastern arm of the ancient Mesozoic Tethyan Seaway (Liu *et al*. 2018). In the southern Palearctic, this event was followed by high extinction rates and the gradual replacement of a former subtropical flora by continental xerophytic and Mediterranean lineages (Manafzadeh *et al*. 2017). The result

from BDSKY is more ambiguous, pointing to a longer period of negative diversification rates, which spans several tens of millions of years, starting at 43 or 34 Ma. The start of this period corresponds to the global cooling event at the Late Eocene-Early Oligocene boundary, the Terminal Eocene Event (TEE) or Late Eocene-Oligocene Cooling Event (Zachos *et al*. 2001), which in the Northern Hemisphere led to widespread extinction and the replacement of a boreotropical flora by temperate elements (Meseguer *et al*. 2018). The fact that BDSKY cannot narrow down the time interval of negative diversification rates can be considered a weakness of the method. However, as Leslie *et al*. (2012) noted, the major families in the conifer tree exhibit different diversification trajectories, which seem to be related to their geographic distribution. Thus, Southern Hemisphere families Podocarpaceae and Araucariaceae comprise on average older (Miocene) clades, and their LTT plots show plateaus (interpreted as signaling an MEE, Harvey *et al*. 1994), extending between 50 and c. 30-25 Ma (cf. Figure 2A in Leslie *et al*. 2012). This signal could correspond with the TEE, which restricted evergreen plant lineages to the equatorial latitudes of the Northern and Southern Hemispheres (Meseguer *et al*. 2018). In contrast, major clades within the Northern Hemisphere Pinaceae and Cupressoideae are mainly of Late Neogene age (< 20 Ma), with the species-rich conifer genera *Pinus, Juniperus* and *Cupressus* diversifying within the last 5-3 Ma (cf. Figure 1D, Leslie *et al*. 2012). The recovery phase detected by BDSKY, with a return to positive diversification rates at c. 3 Ma, could well correspond to this period of rapid diversification reported by Leslie *et al*. (2012). Overall, the ambiguity in the BDSKY results could actually reflect the different diversification trajectories, and their reaction to these sequential MEEs, taken by the conifer families depending on their geographic distribution. Northern Hemisphere conifers in Pinaceae and Cupressoideae were probably as affected by the TEE as the Southern Araucariaceae or Podocarpaceae, but later major climatic events such as the Late Oligocene Warming Event could have obscured the signal of these more ancient events.

Finally, both CoMET and BDSKY (and *TreePar*) assume that changes in speciation and extinction rates occur simultaneously across all branches in the phylogeny ("tree-wide rate shifts"). There is, however, increasing evidence that speciation and extinction rates vary across clades, dependent on biotic factors such as the appearance of key innovation or the colonization of a new ecological niche (Donoghue and Sanderson 2015). Laurent *et al.* (2015) found that *TreePar* suffers from a higher frequency of Type I error under scenarios with substantial clade-rate heterogeneity, whereas CoMET is relatively insensitive to this bias (May *et al*. 2016). It would be interesting to test

the behavior of BDSKY under such scenarios. Another limitation of CoMET or BDSKY is that they model mass extinction events as a non-selective (statistically random) "field of bullets" scenario, in which all lineages in the tree have the same probability of being affected by the mass extinction event (Raup 1982; Harvey *et al.* 1994). Some authors conceive mass extinction as a “fair game” (Darwin-like) scenario, in which the best-adapted species would have the highest survival probability (Raup 1982; Pyron and Burbrink 2012). The geographic distribution of a clade might also condition the impact of the MEE. This could be the case of the conifers, i.e., the older southern lineages exhibit lower turnover rates than the northern lineages, and might have survived the climatic shifts of the Late Cenozoic in the relatively warm or wet oceanic climates of the austral landmasses (Leslie *et al*. 2012). Future method development should consider allowing the magnitude or survival probability to the mass extinction event to vary across clades within the phylogeny.

CONCLUSIONS

Here, we demonstrate that estimating changes in extinction rates through time as in the BDSKY model allows detecting the signature of mass extinction events from phylogenies with only extant taxa. The advantages of a time-slice approach are its closer resemblance to the paleontological record and the possibility to cover a broader range of MEEs, from nearly instantaneous events to a longer time period of elevated extinction rates. However, further simulations are needed to understand the limits of this approach.

Acknowledgments: We thank three anonymous reviewers and the Associate Editor for their insightful comments. V.C. was funded by Ministerio de Economía y Competitividad MINECO Ph.D. (FPI) Fellowship BES-2013-065389, supervised by I.S. I.S. was funded by the Spanish Government and European Regional Development Fund, grant CGL2015-67849-P (MINECO/FEDER) and by Fundacion BBVA “Ayudas a Equipos de investigacion Ecología y Biología de la Conservacion,” Project G999088Q. T.S. is supported in part by the European Research Council under the 7th Framework Programme of the European Commission (PhyPD grant agreement number 335529). The authors also acknowledge the MINECO program “Ayudas de movilidad FPI” for funding a 4-month visit of V.C. to ETH Zürich, Basel, with T.S., which was the inspiration for this project.

Author contributions: V.C., T.S., and I.S. conceived and designed the study; V.C. performed the analyses with help from T.S. and I.S.; V.C., T.S., and I.S. wrote the paper.

Data archiving: All data results from this study are deposited in the public repository

Dryad, under number https://doi.org/10.5061/dryad.qv10c62.

REFERENCES

Alfaro, M. E., Santini, F., Brock, C., Alamillo, H., Dornburg, A., Rabosky, D. L., Carnevale, G. and Harmon, L. J. 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. Proceedings of the National Academy of Science USA 106:13410-13414;

Barnosky, A. D., Matzke, N., Tomiya, S., Wogan, G. O. U., Swartz, B., Quental, T. B., Marshall, C., McGuire, J. L., Lindsey, E. L., Maguire, K. C., Mersey, B. and Ferrer, E. A. 2011a. Has the Earth's sixth mass extinction already arrived? Nature 471:51-57;

Barnosky, A. D., Carrasco, M. A. and Graham, R. W. 2011b. Collateral mammal diversity loss associated with late Quaternary megafaunal extinctions and implications for the future. Geological Society, London, Special Publications 358:179-189;

Beaulieu, J. M. and O'Meara, B. C. 2015. Extinction can be estimated from moderately sized molecular phylogenies. Evolution 69:1036–1043;

Benton, M. J. 2009. The Red Queen and the Court Jester: Species Diversity and the Role of Biotic and Abiotic Factors Through Time. Science 323:728-732;

Boskova, V., Bonhoeffer, S. and Stadler, T. 2014. Inference of Epidemiological Dynamics Based on Simulated Phylogenies Using Birth-Death and Coalescent Models. PLOS Computational Biology 10: e1003913;

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C., Xie, D., Suchard, M. A., Rambaut, A. and Drummond A. J. 2014. BEAST2: A Software Platform for Bayesian Evolutionary Analysis. PLOS Computational Biology 10:doi:10.1371/journal.pcbi.1003537;

Condamine, F. L., Rolland, J. and Morlon, H. 2013. Macroevolutionary perspectives to environmental changes. Ecology Letters: doi:10.1111/ele.12062;

Donoghue, M. J. and Sanderson, M. J. 2015. Confluence, synnovation, and depauperons in plant diversification. New Phytologist 207:260-274;

Ezard, T. H. G., Aze, T., Pearson, P. N. and Purvis, A. 2011. Interplay Between Changing Climate and Species' Ecology Drives Macroevolutionary Dynamics. Science 332:349–351;

Foote, M., Hunter, J. P., Janis, C. M. and Sepkoski Jr., J. J. 1999. Evolutionary and preservational constraints on origins of biologic groups: divergence times of eutherian mammals. Science 283:1310-1314;

Gernhard, T. 2008. New Analytic Results for Speciation Times in Neutral Models. Bulletin of Mathematical Biology 70:1082-1097.

Gould, S. J. 1994. Tempo and mode in the macroevolutionary reconstruction of Darwinism", PNAS 91: 6764–6771;

Harvey, P. H., May, R. M. and Nee, S. 1994. Phylogenies Without Fossils. Evolution 48: 523–529;

Heath, T. A., Huelsenbeck, J. P. and Stadler, T. 2014. The fossilized birth-death process for coherent calibration of divergence-time estimates. Proceedings of the National Academy of Science USA 111: E2957-E2966;
Evol. 28:2577–2589.

Höhna S., Stadler T., Ronquist F., Britton T. 2011. Inferring speciation and extinction rates under different sampling schemes. Molecular Biology and Evolution 28:2577-2589;

Höhna, S. 2014. Likelihood inference of non-constant diversification rates with incomplete taxon sampling. PLOS ONE 9:e84184;

Höhna, S., May, M. R. and Moore, B. R. 2015. TESS: An R package for efficiently simulating phylogenetic trees and performing Bayesian inference of lineage diversification rates. Bioinformatics 32:789-791;

Huges, C. E. and Eastwood, R. 2006. Island ratification on a continental scale: exceptional rates of plant diversification after uplift of the Andes. Proceedings of the National Academy of Science USA 103:10334-10339;

IPCC 2001. Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental

Panel on Climate Change, Cambridge University Press;

IUCN. International Union for Conservation of Nature Red List of Threatened Species. www.iucnredlist.org

Jablonski, D. 2008. Extinction and spatial dynamics of biodiversity. Proceedings of the National Academy of Science USA 105:11528-11535.

Laurent, S. S. J., Robinson-Rechavi, M. and Salamin, N. 2015. Are we able to detect mass extinction events using phylogenies? BMC Evolutionary Biology doi: 10.1186/s/s12862-015-0432-z;

Leslie, A. B., Beautlieu, J. M., Rai, H. S., Crane, P. R., Donoghue, M. J. and Mathews, S. 2012 Hemishere-scale differences in conifer evolutionary dynamics. Proceedings of the National Academy of Science USA 109:16217-16221.

Liu, H., Li, S., Ugolini, A., Momtazi, F., & Hou, Z. 2018. Tethyan closure drove tropical marine biodiversity: Vicariant diversification of intertidal crustaceans. Journal of Biogeography, 45:941-951.

May, M. R., Höhna, S. and Moore, B. R. 2015. A Bayesian approach for detecting mass-extinction events when rates of lineage diversification vary. bioRxiv preprint first posted online May 31, 2015; doi: http://dx.doi.org/10.1101/020149.

May, M. R., Höhna, S. and Moore, B. R. 2016. A Bayesian approach for detecting the impact of mass-extinction events on molecular phylogenies when rates of lineage diversification may vary. Methods in Ecology and Evolution 7: 947-959.

Manafzadeh, S., Staedler, Y. M. and Conti, E. 2017. Visions of the past and dreams of the future in the Orient: The Irano-Turanian region from classical botany to evolutionary studies. Biological Reviews 92:1365-1388.

Meseguer, A. S. Lobo, J. M., Cournault, J., Beerling, D., Rufhel, B, Davis, C., Josselin, E. and Sanmartín, I. 2018. Reconstructing deep-time paleoclimate legacies in the clusioid Malpighiales unveils their role in the evolution and extinction of the boreotropical flora. Global Biogeography and Ecology 7: 616-628.

Purvis, A. 2008. Phylogenetic Approaches to the Study of Extinction. Annual Review of Ecology, Evolution, and Systematics 39:301-319;

Pyron, R. Alexander, and Burbrink, F. T. 2012 Extinction, ecological opportunity, and the origins of global snake diversity. Evolution 66: 163-178.

Rabosky, D. L. 2006. Likelihood methods for detecting temporal shifts in diversification rates. Evolution 60:1152-1164;

Rabosky, D. L. 2014. Automatic Detection of Key Innovations, Rate Shifts, and Diversity-Dependence on Phylogenetic Trees. PLOS ONE 9: e89543;

Rannala, B. 2002. Identifiably of parameters in MCMC Bayesian inference of phylogeny. Systematic Biology 51:754-760;

Raup, D. M. 1979. Size of the Permo-Triassic Bottleneck and Its Evolutionary Implications. Science 206: 217-218;

Raup, D. M. 1982. Extinction: bad genes or bad luck? ACTA GEOLÒGICA HISPÀNICA. Concept and method in Palaeontology 16: 25-33;

Sanmartín, I. and Meseguer, A. S. 2016. Extinction from Phylogenetics and Biogeography: From Timetrees to Patterns of Biotic Assemblage. Frontiers in Genetics 7: doi: 10.3389/fgene.2016.00035;

Sepkoski Jr., J. J. 1982. Mass extinctions in the Phanerozoic oceans: A review. Geological Society of America Special Papers 190: 283–290;

Stadler, T. 2011a. Inferring speciation and extinction processes from extant species data. Proceedings of the National Academy of Science USA 108:16145-16146;

Stadler, T. 2011b. Mammalian phylogeny reveals recent diversification rate shifts. Proceedings of the National Academy of Science USA 108: 6187–6192;

Stadler, T. 2011c. Simulating trees with a fixed number of extant species. Systematic Biology 60:676–684;

Stadler T. and Bokma, F. 2012. Estimating Speciation and Extinction Rates for Phylogenies of Higher Taxa. Systematic Biology 62:220-230;

Stadler, T., Kühnert, D., Bonhoeffer, S. and Drummond, A. J. 2013. Birth–death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). Proceedings of

the National Academy of Science USA 110:228-233.

Valente, L. M., Savalainen, V. and Vargas, P. 2010. Unparalleled rates of species diversification in Europe. Proceedings of the Royal Society B: Biological Sciences 277:1489–1496;

Wiens, J. J., Ackerly, D. D., Allen, A. P., Anacker, B. L., Buckley, L. B., Cornwell, H. V., Damschen, E. I., Davies, T. J., Grytnes, J.-A., Harrison, S. P., Hawkins, B. A., Holt, R. D., McClain, C. M. and Stephens, P. R. 2010. Niche conservatism as an emerging principle in ecology and conservation biology. Ecology Letters 13:1310-1324.

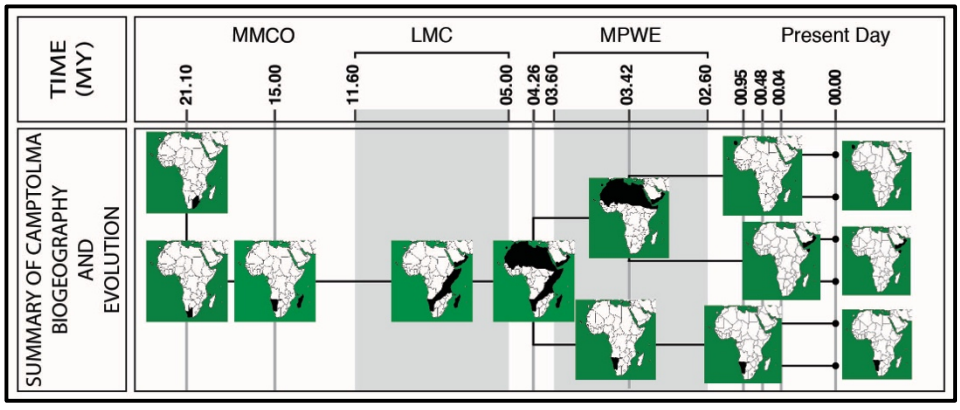
Zachos, J., Pagani, M., Sloan, L., Thomas, E. and Billups, K. 2001. Trends, Rhythms, and Aberrations in Global Climate 65 Ma to Present. Science 292: 686-693.

CHAPTER 2

Combining Bayesian biogeographic inference and phylogenetically-informed niche models to reconstruct the role of ancient climate change in depauperate lineages

This chapter has been realised in collaboration with Isabel Sanmartín (Department of Biodiversity and Conservation, Real Jardín Botánico, CSIC, Spain) and Mario Mairal (Department of Botany and Zoology, Stellenbosch University, South Africa).

This chapter has been submitted to the peer-review journal *Systematic Biology* on 26th September 2019.



CHAPTER 2

Combining Bayesian Biogeographic Inference and Phylogenetically-informed Niche Models to Reconstruct the Role of Ancient Climate Change in Depauperate Lineages

Victoria Culshaw, Mario Mairal, Isabel Sanmartín

ABSTRACT

The current aridification that affects the Mediterranean region and northern Africa, a change from wetter to drier climates with increasing annual temperatures and decreasing precipitation, is a major scientific and societal concern. This trend is not recent but has been ongoing for the last 25 million years, driven by global and regional geotectonism. Reconstructing the evolutionary and ecological responses of plant lineages to long-term aridification could help increase the accuracy of forecast predictions under scenarios of global climate warming. Two approaches are often used to understand the evolutionary signature of climate change on species range dynamics: i) ecological niche models (ENM) estimate the environmental preferences of a species based on occurrence data, which are then projected over palaeoclimate layers to explore for similar conditions in the past; ii) in biogeographic inference, a time-calibrated phylogeny with species distributions is used to infer ancestral ranges and past events of geographical movement. There have been attempts to integrate both approaches, but they typically require large datasets. Here, we explore a framework to combine them as independent but complementary sources of evidence for inferring evolutionary history in depauperate lineages with restricted distributions and limited associated data. We use as a case study, genus *Camptoloma*, a genus with only three species showing one of the largest known intracontinental disjunctions, between Macaronesia, Eastern Africa/Southern Arabia and Southwest Africa, in what is termed the African Rand Flora pattern. Using Bayesian biogeographic inference based on nuclear and chloroplast DNA markers, combined with past and present ENM geographic projections calibrated with phylogenetic/biogeographic data, we show that the current disjunct distribution of *Camptoloma* across Africa was likely the result of fragmentation and extinction/population bottlenecking events associated to historical aridification cycles, in line with the "climatic refugia" hypothesis. We also present an approach to use evolutionary data (phylogenetic and biogeographic information) for selecting the truncation threshold in ENMs.

Running-title: Biogeographic Niche Model in Rand Flora Camptoloma

Keywords: Ancient Climate Change, Bayesian Inference, Biogeographic Reconstruction, Ecological Niche Model, Extinction, Truncation Threshold, Macroevolution, Rand Flora

INTRODUCTION

The current concern on anthropogenic-induced climate change and its impact on biodiversity levels has increased the interest in reconstructing organism responses to past climatic events, including adaptation (evolving a new trait), persistence in their geographical locations (phenotypic plasticity), geographic range shifts (displacement, expansion and contraction), and extinctions (Diniz-Filho and Binin 2008; Willis and McDonald 2011;

Araújo et al. 2013; Romdal et al. 2013; Meseguer et al. 2018). While numerous studies have focused on changes in species distributions under scenarios of future climate change using extrapolative ecological niche models and present-day occurrence data (e.g. Carboni et al. 2018; Hæuser et al. 2018), the current expectation is that including long-term evolutionary data will help increase the accuracy of our forecast predictions (Martínez-Meyer et al. 2004; Diniz-Filho and Binin 2008; Romdal et al. 2013; Meseguer et al. 2015; Burke et al. 2018).

Geographical range shifts are often predicted as the outcome of rapid climate change, under the assumption that the rate of environmental change is greater than that of a species adaption (Martínez-Meyer and Peterson 2006; Thuiller et al. 2006; Waldron 2010). During climatically adverse periods, species may survive by expanding their geographic range when tracking their niche ("climatic refugia"), from which they disperse, once conditions become more favourable, hence expanding their geographic range. These range contractions can result in smaller population sizes, reduced genetic diversity and a higher extinction risk (Waldron 2010; Mairal et al. 2018). However, a change in the physical template that is too large for populations to migrate/adapt may result in fragmentation of the species' original distribution into smaller, spatially disjunct ranges (Wiens 2004). Reproductive isolation and reduced gene flow among these disconnected ranges might lead to larger inter-population genetic differences and eventually into speciation (Dorn et al. 2014).

A prime example of this scenario is the glacial and interglacial stages of the Pleistocene and the Holocene in the last 2.6 millions of years (Hewitt 2004). Many studies have focused on the signature of these geologically recent climatic events in within-species population dynamics (Hewitt 2004; Martínez-Meyer et al. 2004; Espíndola et al. 2012; Martínez-Meyer and Peterson, 2006) as a means to understand population responses to current and future climate change (Mairal et al. 2018). These short temporal scales, however, might be limited in their capacity to predict evolutionary responses to climate warming, especially in the role of persistence and adaptation, which is built over generations of genomic changes (Svenning et al. 2015). Recent studies integrating present-day and fossil occurrence data have found that under scenarios of past global climate change, similar in intensity to the current one (e.g. the Early Eocene Climatic Optimum, EECO, 53-50 million years (Ma) or the Mid Pliocene Warming Event, MPWE, 3.6 Ma, Zachos et al. 2008), persistence and adaptation was a more frequent response in plant lineages than geographical range shifts (Willis and McDonald 2011; Burke et al. 2018; Meseguer et al. 2018). Complementary studies on species range dynamics spanning time periods of several million years may be required in order

to recover the evolutionary signature of climate change (Svenning et al. 2015).

Two approaches are often used to understand the evolutionary signature of climate change on species range dynamics (Svenning et al. 2015). First, phylogenetic and geographic data, e.g. a time-calibrated phylogeny with species distributions, can be used to infer ancestral ranges and events of geographical movement using biogeographic inference methods (Ronquist and Sanmartín 2011; Mairal et al. 2015a). Second, Ecological Niche Models (ENM) can be used to estimate the environmental preferences (tolerance range) of a species based on its current distributional range, which are then projected over palaeoclimate scenarios to explore spatially and temporally for similar conditions in the past (Araújo and New 2007; Meseguer et al. 2015; Mairal et al. 2017). The two approaches have their strengths and drawbacks. Biogeographic inference is often based on user-defined areas that lack geographical meaning (but see Tagliacollo et al. 2015). Incomplete or geographically biased taxon sampling and high extinction rates (especially if area-dependent, Meseguer et al. 2015) might mislead biogeographic reconstructions through the selective removal of speciation events from the phylogeny (Sanmartín and Meseguer 2016). Conversely, ENMs use geographically meaningful coordinates of presence-only records as biological input, but they are typically limited by the availability of palaeoclimate data, having been mostly applied to recent geological times (Martínez-Meyer et al. 2004, Mairal et al. 2018; but see Meseguer et al. 2015). Also, ENMs implicitly assume that species climatic tolerances have been conserved over evolutionary time, which might be unrealistic under repeated cycles of climatic change and long geological time periods (Peterson 2006). Recent studies have shown that integrating information from the fossil record can increase the reliability of biogeographic reconstructions models and ENMs (Metcalf et al. 2014; Meseguer et al. 2015). Integrating the two approaches, i.e. using

palaeoclimate ENM projections to add areas outside the current distribution and inform dispersal models, has also been shown to improve the accuracy and realism of biogeographic reconstructions (Smith and Donoghue 2010; Meseguer et al. 2015).

The current aridification process that affects the Mediterranean region and northern Africa, a change from wetter to drier climates with increasing annual temperatures and decreasing precipitation, is a major concern under a scenario of global climate warming (Giorgi and Lionello 2008). This trend is not recent but has been ongoing for the last 25 million years, driven by global and regional geotectonis events (Liu et al. 2018). In Africa, this led to the appearance of semi-arid and desert climates in northern and southwestern Africa and the expansion of savannahs in the

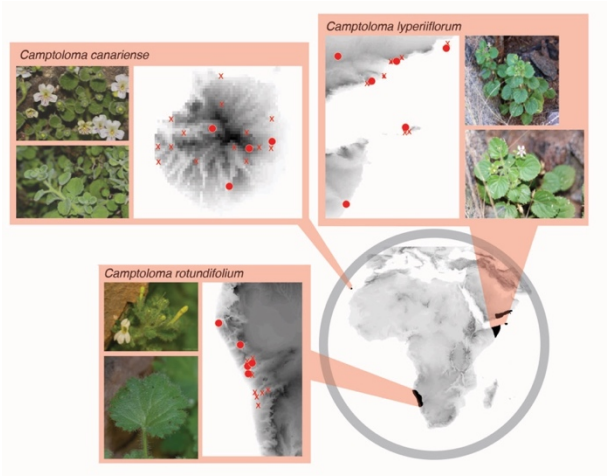


Figure 1. Geographic distribution of genus *Camptoloma*, showing the east-west-south disjunction among the three species, spanning thousands of kilometres across the African continent. Red crosses represent reliable location geographical coordinates from GBIF, while red dots represent locations recorded from herbarium vouchers.

east (Trauth et al. 2009; Senut et al. 2009). Phylogenetic and fossil data suggest that plant genera responded in different ways to these aridification events, from extinction and geographic range shifts in tropical and temperate-adapted lineages (Pokorny et al. 2015), to rapid diversification driven by allopatry and adaptation in other groups (Thiv et al. 2010; García-Aloy et al. 2017; Pirie et al. 2018).

The Rand Flora (RF) pattern is a continental-scale geographic disjunction in which sister species are distributed on opposite sides of the African continent (Sanmartín et al. 2010). The classical example is the east-west disjunction between Macaronesia and Eastern Africa observed in genus *Canarina* (Campanulaceae, Mairal et al. 2015a) (Supplementary Information Fig. S1). Molecular estimates have dated RF disjunctions from the Early Miocene to the Pleistocene, though most trace back to the Mid-Late Miocene and Pliocene periods, concurrent with the intensification of the aridity trend (Pokorny et al. 2015). RF lineages belong to different angiosperm families, with dissimilar morphology and life forms, but they typically exhibit subtropical, temperate or semi-arid affinities, i.e. they do not occur in deserts (Sahara) or in the tropical lowlands (Guinea-Congo Basin). Species populations are often small in size and have highly restricted distributions (Pokorny et al. 2015; Mairal et al. 2017a, 2018). As in many African plant groups, there is little data on their reproductive biology or evolutionary history, and distributional records are scarce (www.gbif.org, (9th July 2014) GBIF Occurrence Download https://doi.org/10.15468/dl.abs9dk), especially for difficult-to-access or politically unstable regions (e.g. Somalia), though there have been recent efforts to address this (Mairal et al. 2017a; Villaverde et al. 2018). Moreover, with a few exceptions (e.g. Thiv et al. 2010), RF phylogenies comprise less than 5-10 species and often exhibit "broom and handle" shapes, with long stem-branches and young crown clades, indicative of high extinction rates (Antonelli and Sanmartín 2011; Mairal et al. 2015a; Sanmartín and Meseguer 2016).

One of the most extreme examples of the RF disjunction is genus *Camptoloma*

(Scrophulariaceae, with three isolated species in Macaronesia (*Camptoloma canariensis* in Gran Canaria), Eastern Africa (*C. lyperiiflorum* occurring in Somalia, Yemen, Oman and the Socotra Archipelago), and southwest Africa (*C. rotundifolium* in Namibia), shown in Figure 1. Species populations exhibit restricted geographic ranges, occupying humid microclimates within a more arid geographical template (Domina et al. 2012; Craven and Craven 2000). Phylogenetic relationships among species and allies remain controversial due to low phylogenetic support and limited sampling; i.e. existing molecular phylogenies have been based on three chloroplast DNA (cpDNA) markers (*ndhF*, *trnLF*, and *rps16*), with species represented by one specimen each (Kornhall et al. 2001; Oxelman et al. 2005; Pokorny et al. 2015). Pokorny et al. (2015) dated the origin of RF disjunction of *Camptoloma* (crown-node *Camptoloma*) in the Miocene-Pliocene boundary (5.5 Ma), whereas the divergence with its sister group (South African *Phygellius capensis*) was estimated much older, in the Early Miocene (21 Ma). These deep Miocene divergences, "broom and handle shaped" phylogeny, and microrefugia-type distributions makes *Camptoloma* an ideal group to test the hypothesis of RF disjunctions as climatic relicts linked to African aridification history. Yet, the small-sized phylogeny (3 tips), large temporal gaps, and the lack of fossil record in RF lineages (Sanmartín et al. 2010) hamper the use of sophisticate diversification and ancestral reconstruction methods for inferring evolutionary history (Sanmartín and Meseguer 2016). Representation of *Camptoloma* records on online databases and herbaria is also limited and likely biased due to the species occurrence in difficult-to-access or politically unstable regions (e.g. GBIF contains 35 records of *C. lyperiiflorum* in Yemen, 7 in Oman but none in Somalia). This is problematic when using statistical methods for integrating ENMs and biogeographic inference, as done in other RF lineages with larger fossil occurrence datasets (Meseguer et al. 2015, 2018).

Here, we explore a novel framework to combine these two approaches, ENM and biogeographic inference, as independent but complementary sources of evidence to infer the spatiotemporal evolution of species-poor lineages with restricted distributions and limited associated data. First, we inferred phylogenetic relationships and lineage divergence times in *Camptoloma* based on a large dataset of infraspecific sequences and representative Scrophulariaceae outgroups for seven noncoding cpDNA markers and the nuclear ribosomal ITS region. Second, we reconstructed ancestral ranges and the history of geographic range shifts using hierarchical Bayesian Inference approaches (BI). BI methods are ideally suited for analysis with limited data because they allow marginalising over the variation (uncertainty) in nuisance parameters such as molecular states, when

computing the posterior probability distributions of rates of range evolution and geographical states (Sanmartín et al. 2008). At the genera and species level in Scrophulariaceae, we used a Bayesian approach to Dispersal-Extinction-Cladogenesis (Ree and Smith 2008) implemented in the free software RevBayes (Höhna et al. 2016; Landis et al. 2018). At the population level within individual *Camptoloma* species, we employed Discrete Trait Analysis (Lemey et al. 2009) and the new structured coalescent model BASTA (De Maio et al. 2015) implemented in BEAST (Bouckaert et al. 2014). Third, we used ENMs hindcasted upon palaeoclimatic layers representing global warming or cooling events from the Early Miocene to the present day to predict changes in species and population geographic ranges over time. To compensate for using limited and potentially biased palaeoclimatic data, and an occurrence only dataset in an ENM that required presence/absence data (i.e. the issue of modelling pseudoabsences, Chefaoui and Lobo 2008), we used an approach that combines the advantages of point pattern analysis, habitat suitability distances (environmental niche factor analysis, ENFA) and logistic regression-kriging (Hengl et al. 2009). Rather than integrating BI and ENM approaches into one analysis to reconstruct spatiotemporal evolution (Meseguer et al. 2015), we used them as independent sources of evidence that complement each other in regard to data limitations. For example, the long branch of c. 15 million years spanning the divergence between stem- and crown-node *Camptoloma* (Pokorny et al. 2015) introduces significant uncertainty in ancestral state reconstruction (Sanmartín and Meseguer 2016). Here, this is reduced by drawing information from hindcasted ENM palaeoclimate projections for those geological time periods that lie in-between. Similarly, instead of using a subset of the original dataset (Chefaoui and Lobo 2008), we propose a novel approach in which phylogenetic and biogeographic information are used to provide truncation thresholds for defining habitat suitability in ENM models for past, present, and future climate change scenarios.

Specifically, our goals were to: (i) resolve phylogenetic relationships among the three species of genus *Camptoloma* and its systematic position within Scrophulariaceae; (ii) combine information from Bayesian biogeographic inference and palaeoclimatic hindcasted ENMs to infer the origins of the RF disjunction; (iii) forecast the fate of *Camptoloma* under scenarios of future climate warming; and finally (iv) propose an alternative way to infer evolutionary history in depauperate lineages with limited data, which might be the result of high extinction rates linked to historical aridification cycles.

MATERIAL AND METHODS

Phylogenetic and Biogeographic Inference

Taxon sampling and DNA sequencing: DNA was extracted from silica gel-dried leaves (*Camptoloma canariense*) obtained through field expeditions to Gran Canaria (2009-2014), and from loans of dried material (*C. lyperiiflorum, C. rotundifolium*) from different herbaria (Real Jardín Botánico (MA), Royal Botanical Garden of Edinburgh (E), Uppsala Museum of Evolution Herbarium (U) and Pretoria National Herbarium (PRE)). In total, we collected 35 individuals representing 15 *Camptoloma* populations, which span the genus' distributional range: 8 individuals of *C. canariense*, 17 of *C. lyperiiflorum* and 10 of *C. rotundifolium*. Additionally, 24 representatives of several tribes of Scrophulariaceae were included as outgroups: Myoporeae, Leucophylleae, Aposimeae, Scrophularieae, Limoselleae, Buddlejeae, and Teedieae. Where possible, species and genera were represented by several individuals. In total, our dataset comprised 73 samples. Taxon names, location, voucher information, and GenBank accession numbers are provided in Supplementary Information Table S1.

DNA was extracted using the DNeasy Plant Mini Kit (QIAGEN Inc., California, USA) according to manufacturer's instructions. Sequences for seven non-coding cpDNA regions –the intergenic spacers *trn*L-*trn*F, *trn*S-*trn*G, *rpl*32-*ndh*F, *psb*J-*pet*A, *pet*B-*pet*D, *trn*T-*trn*L and the *rps16* intron– and the multicopy nuclear marker ITS were obtained using universal plant primers and newly-designed primers for difficult herbarium specimens. See SI "Extended Material and Methods" and Table S2 for more details on primers and specific PCR protocols. Several specimens failed for a few markers, especially for the outgroup taxa. When available, for these we used equivalent sequences from the same marker and species obtained from GenBank (e.g. *Hebenstretia dentata*: *trnL-trnF*; *Plantago*: ITS, *rpl32-ndhF, pet*B-*pet*D, *rps16* intron, *trn*L-*trn*F, *trnS-trnG*; see Table S1 for accession numbers). In all, we generated 521 new sequences from 73 specimens: ITS (71 sequences), *rpl*32-*ndh*F (73 sequences), *pet*B-*pet*D (59 sequences), *psb*J-*pet*A (62 sequences), *rps*16 intron (67 sequences), *trn*L-*trn*F (64 sequences), *trn*S-*trn*G (62 sequences) and *trn*T-*trn*L (66 sequences).

Phylogenetic analyses: Sequences were aligned using MUSCLE v3.8.31 (Edgar 2004; www.drive5.com/muscle) with a maximum of eight iterations and adjusted manually in Geneious Pro 5.6.7 (www.geneious.com). We used three datasets for the analyses: (i) the "all specimens" dataset included all sequences from our study; (ii) the "outgroup" dataset (44 sequences) included all outgroup taxa and two specimens (representing different populations) for each species of *Camptoloma*; (iii) the three "population-level"

datasets included all sequences in each individual *Camptoloma* species: "C. *canariense*" (64 sequences), "*C. lyperiiflorum*" (121 sequences) and "*C. rotundifolium*" (66 sequences). The "all specimens" dataset consisted of 521 sequences with 7944 nucleotide sites, while the combined "population-level" *Camptoloma* datasets consisted of 247 sequences with 6626 nucleotide sites. Table S3 summarises some statistics of the genomic regions studied for the different datasets.

Phylogenetic inference was performed in a MCMC Bayesian framework using MrBayes v3.2.6 (Ronquist et al. 2012). The script to run this analysis is provided in Appendix A1. Each MrBayes analysis was run for 10,000,000 generations, sampling every 1000th generation, with four chains in two parallel searches. Convergence and effective mixing was assessed by ensuring that the effective sample size (ESS) for each parameter reached 200 in Tracer v1.7 (Rambaut et al. 2018) and the Potential Scale Reduction Factor (PSRF) approached in MrBayes; differences in topology among runs were assessed through the split frequencies in MrBayes (< 0.1). The posterior probability tree distribution was summarised in a 50% majority-rule consensus tree with 95% credibility intervals, after removing 25% of samples (burn-in), and visualised in FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree).

We used the "all specimens" dataset for the phylogenetic inference. Each marker was first analysed separately to assess congruence among the resultant tree topologies. The software jModelTest v.2.1.10 (Darriba et al. 2012) was used to select the best-fit nucleotide substitution model (results presented in Appendix 2); some of these models are not implemented in MrBayes (e.g. TVM+G) and were replaced by the most similar model: GTR+I+G was selected for ITS and GTR+G for all cpDNA markers. Comparison among individual consensus trees revealed little resolution and no incongruent clades receiving significant Bayesian support (> 95%); results not shown. Additionally, we carried out a sensitivity analysis based on Bayes Factor comparisons to decide the best partitioning strategy for a concatenate nuclear-chloroplast dataset: 1) nuclear marker (ITS) only; 2) concatenate cpDNA dataset partitioned by marker; 3) concatenate nuclear-cpDNA dataset partitioned by genome (ITS vs. cpDNA); and 4) concatenate nuclear-cpDNA dataset partitioned by marker. Indel (gap) information was included in each aligned matrix using the Simmons and Ochoterena (2000) simple coding algorithm in SeqState (Müller 2005); this was modelled as a single partition under the restriction site (F81) model according to the MrBayes manual. The selected model was a concatenate dataset with partition strategy 3, with substitution models GTR+I+G for ITS,

GTR+G for the cpDNA partition and F81 for indels. See SI Extended M&M for more details.

***Divergence time estimation**:* Lineage divergence times were estimated using Bayesian relaxed molecular clock models implemented in BEAST v1.8.2 (Drummond et al. 2012). The script to run this analysis is provided in Appendix A3. First, a higher genus-level analysis was run to estimate the crown-age of *Camptoloma* using the *"all specimens"* dataset. As there are no known fossils of Scrophulariaceae, we used secondary calibration estimates from Magallón et al. (2015)'s fossil-calibrated angiosperm analysis. We calibrated the following nodes using normal distribution priors spanning the 95% High Posterior Density (HPD) credibility intervals obtained by Magallón et al.: (i) the divergence between Scrophulariaceae and sister-family Plantaginaceae (stem-node Scrophulariaceae): mean = 48.6 Ma, 95% HPD = 37.9-62.5, here represented by genus *Plantago* (see Table S1): (ii) the crown-node of Scrophulariaceae (mean = 42.12 Ma, 95% HPD = 29.5-58.5 Ma); and (iii) the split between *Scrophularia* and *Verbascum* (mean = 19 Ma, HPD = 8.1-36.4 Ma). The uncorrelated lognormal distribution (UCLD) was used as the molecular clock and the birth-death model with incomplete sampling as the tree prior (Stadler 2009); substitution models were unlinked following the partitioning strategy see above. Convergence and mixing were assessed using Tracer. A maximum clade credibility tree (MCC) was constructed in TreeAnnotator v. 1.8.2 (Drummond et al. 2012) after removing 25% posterior trees as burn in. See SI Extended M&M for more details.

To account for potential biases introduced by the single branching process assumption when combining data from different evolutionary levels (populations, species and genera), we used an alternative "nested-dating" approach (Pokorny et al. 2011; Mairal et al. 2015a). The script to run this analysis is provided in Appendix A4. In this approach, the family-level "outgroup" dataset was linked through the molecular clock with the three "population-level" *Camptoloma* datasets. The first was modelled under a birth-death tree prior, while the Camptoloma "population" datasets were assigned unlinked, constant-size, coalescent priors. We ran analyses under the strict and UCLD clock models, though the latter did not converge and was discarded. All other parameters followed the settings as above.

Biogeographic and phylogeographic analyses: Table S4 lists the taxa included in the biogeographic analysis and their geographic distribution, grouped by continental region. Distribution data for each species and genera was obtained from online open source databases (e.g. GBIF); online plant guidebooks (e.g. http://southernafricanplants.net and https://species.nbnatlas.org), and herbarium

data for our DNA vouchers. In some cases, the geographic range of the genus was not fully covered by that of the representative species included in the analysis, e.g. for the widespread genera *Myoporum* and *Buddleja* (Table S4).

To reconstruct the biogeographic history of *Camptoloma* in relation to its allies within Scrophulariaceae, we used the Dispersal-Extinction-Cladogenesis model (DEC, Ree and Smith 2008), implemented in a Bayesian framework in RevBayes (Höhna et al. 2016) as described in Landis et al. (2018). Using Bayesian inference allows incorporating the uncertainty in biogeographic parameter estimation through the computing of marginal posterior probabilities for ancestral ranges and rates. We ran two analyses of 10,000 generations, sampling every 10^{th} generation on the MCC tree obtained from the nested-dating approach, pruned to include one individual per species. The dispersal rate was modelled with a lognormal prior between 10^{-4} and 10^{-1} events per million years, whereas all other priors were set as default (https://revbayes.github.io/tutorials/biogeo/biogeo_intro.html). Cladogenetic range evolution was input as allopatry (vicariance or peripheral isolate speciation; Ree et al. 2005) for widespread taxa and subset sympatry for single areas, using a simplex (1, 1) prior. We did not use the DEC+J model allowing for jump dispersal ("j" parameter) because of recent criticisms to this model (Ree and Sanmartín 2018). The script to run this analysis is provided in Appendix A5. Six operational areas were defined based on distribution patterns in *Camptoloma* and closest genera: (A) Gran Canaria, (B) Namibia, (C) western half of South Africa, (D) eastern half of South Africa, (E) eastern Africa, (F) southern Arabia (Yemen/Oman/Socotra/Somalia), and a seventh area to cover the distribution of the outgroup taxa in the rest of the world named as "Not in Africa", see map in Figure 4. We did this because the number of possible ranges (states) in the DEC model increases exponentially with the number of areas, so that the analysis becomes intractable for more than 10 areas (Landis et al. 2013). Also, outgroup taxa were not well represented in our analysis, and we were mostly interested in migration events within Africa and nearest "Rand Flora" regions.

To infer phylogeographic history at population-level within each *Camptoloma* species, we used the Bayesian Discrete Trait Analysis approach (DTA, Lemey et al. 2009) implemented in BEAST 1.8. This model –akin to the Bayesian Island Biogeography (BIB) model developed by Sanmartín et al. (2008)– uses Bayesian MCMC and a nucleotide evolutionary model to infer the history of migration events between individuals and single DNA copies. For the analyses, we used the "population-level" *Camptoloma* datasets calibrated with the crown-age inferred in the nested-dating analysis. The topology was further constrained to follow the MCC trees

from these analyses. We implemented a symmetric migration model with 21 dispersal rates under a UCLD clock model, and applied Bayesian Stochastic Variable Selection (BSVS) to find the dispersal pathways with non-negligible support (Lemey et al. 2009). Details on other parameter priors can be found in SI Extended M&M and the DTA script can be found in Appendix 6.

De Maio et al. (2015) criticised DTA because of the unrealistic treatment of the migration-mutation process. Since the effect of migration on effective population sizes is not modelled in the likelihood of the coalescent (i.e. the dataset is assumed to belong to a constant-size, panmictic population), posterior probability values on ancestral inferences tend to be overestimated and the method is highly sensitive to unequal sampling effort among areas (De Maio et al. 2015). To assess the effect of this bias, we ran De Maio et al.'s Bayesian Structured Coalescent Approximation (BASTA), implemented in BEAST v2.4.7 (Bouckaert et al. 2014), which uses the structured coalescent process (MTT model, Vaughan et al. 2012) to model migration but is computationally more efficient in handling a large number of populations/areas. We implemented a migration-mutation Volz model with symmetric transition rates and equal population sizes under a UCLD model. See more details in SI Extended M&M and the BASTA script can be found in Appendix 7.

Ecological Niche Modelling

Distributional and climatic data: Records of *Camptoloma* with geographical coordinates were obtained from herbarium sample vouchers, the online location database GBIF and georeferenced populations obtained during fieldtrips to Gran Canaria www.gbif.org, (9th July 2014) GBIF Occurrence Download https://doi.org/10.15468/dl.abs9dk).

Present-day climate data for the African continent, the Canary Islands, Madagascar, Yemen and Oman were obtained from WorldClim2 (http://worldclim.org/version2, Fick and Hijmans 2017) at intervals of 2.5 minutes (grid cells of 20 x 20 km) for two sets of variables (listed in Table S5): (i) monthly measurements of temperature (minimum/maximum/mean), precipitation, solar radiation, water vapour pressure and wind speed, which have been averaged over the years ranging between 1970-2000; and (ii) nineteen bioclimatic (BIO) variables, which are derived from the monthly values of precipitation and mean temperature variables.

All data variables were normalised to lie within the closed boundaries [0-1]:

$$z_i = \frac{\text{variable}_i - \min(\text{variable})}{\max(\text{variable}) - \min(\text{variable})}$$

where z is the i^{th} normalised data variable. This climatic dataset (26 variables in total) was then matched against the geographical coordinates above to generate a "present-day"

dataset for the current distribution of *Camptoloma* in Africa and adjacent regions. We created separate datasets for each species of *Camptoloma* and for the genus as a whole (Appendix 8a).

Ecological Principal Component Analysis: Ecological Principal Component Analysis (e-PCA, implemented with the R function *prcomp*{*stats*} following the methods of Novak et al. (2010)) was used on the *Camptoloma* "present-day" dataset to identify the variables that best explain the climatic tolerances of each species, and to explore the conservation versus segregation of climatic niches among species. If data points in the environmental space defined by the PCA were clustered by species identity, it would suggest that species occur in environmentally well-defined, non-overlapping climatic niches; if no clusters are found, it would support conservation of climatic preferences at the genus level. The same reasoning can be applied to populations within species. For this analysis, we used the mean monthly values of the seven climatic variables, which were averaged over the 12 months to obtain "yearly" values, and the 19 BIO variables (Table S5) only for the locations where *Camptoloma* individuals are present.

Geographical projections of ENMs using phylogenetically informed truncation thresholds: To generate habitat suitability models for *Camptoloma* species over time (past, present and future), we used Hengl et al. (2009) approach to predict a species' potential distribution or "fundamental climatic niche" based on presence-only distribution data. This approach uses R code from the *spatstat*, *adehabitat* and *gstat* packages (code available at spatialanalyst.net /wiki/ index.php/Species_Distribution_Modelling; Hengl et al. 2009) and combines point pattern analysis (kernel smoothing), Environmental Niche Factor Analysis (ENFA), and logistic regression kriging to predict habitat suitability from occurrence-only data. It is an extension of the approach proposed by Engler et al. (2004) and addresses the issues highlighted by subsequent authors (e.g. Chafaoui and Lobo 2007) on the use of pseudoabsences in ecological niche modelling. First, point pattern analysis and ENFA were used to generate habitat suitability values from the present-day full dataset; this information was then used to weigh the selection of pseudo absences to create a "presence/pseudo absence" dataset needed for generating the regression-kriging GLM models used to predict species' habitat suitability across geographic space and hence location (Hengl et al. 2009). In total, 48 present-day ENMs were created, one for each month of the year for each *Camptoloma* species and for the genus as a whole.

A common practice in ecological niche modelling is to truncate the continuous value

predictions returned from ENMs to produce binary representations of presence/absence of individual data (Diniz-Filho et al. 2010). This truncation threshold if often selected using model fitting techniques and a subset of the original data ("training dataset", Diniz-Filho et al. 2010; Allouche et al. 2006). Here, we propose a different approach in which we use information from phylogenetic biogeographic inference to set the truncation threshold in ENMs (described in Fig. 2). First, the results from the DEC biogeographic analysis, nodal ancestral distributions, were represented as "binary" (presence/absence) maps; for some nodes, there could be several possibilities (ancestral ranges) inferred by the RevBayes-DEC analysis (Figs. 2a, 4). Along branches connecting two nodes, alternative ancestral distributions are suggested ("Sugg. Dist.", Fig. 2b). Second, the monthly continuous ENM predictions ("ENM raw data", Fig. 2c) are run through truncation thresholds, with the scale ranging from 0.05 to 0.95 in 0.05 increments. Grid cells with values that are equal to or greater than the truncation threshold are assigned a value of 1, indicating habitat suitability; those that fall below the threshold are given a value of 0, i.e. inhospitable habitat ("ENM Post Threshold", Fig. 2d). Next, the post-truncation ENMs ("ENM Post Threshold", Fig. 2d) are examined to find "clusters" or "connected" suitable habitats for each truncation threshold: suitable cells are considered connected if the distance between them is ≤ two unsuitable cells ("ENM Post Threshold", pink highlights Fig. 2d). Clusters for each month were then overlaid to construct "yearly" (twelve-month) summaries ("Year Summary", Fig. 2e) Finally, yearly summaries for different truncation threshold values are compared to find the one(s) that best match DEC inferred nodal ancestral ranges, and the suggested ancestral ranges along internodes ("Chosen Truncation Threshold", Fig. 2f). For present-day ENMs, the truncation threshold was selected by comparing yearly summaries against the current distribution of each species ("RevBayes Summary", maps shown at tips, Fig. 2a): we selected those truncation threshold values that yielded predicted distributions that were similar but not identical to the current distribution of the extant species.

Present-day ENM projections: Before creating the present-day ENMs, we split the "present-day" dataset by monthly variables and ran a PCA on each (12 PCAs in total) to reduce their dimensions through the removal of between-variable correlations and non-informative variables. Only variables with a factor score ≥ 0.6 or ≤ -0.6 with the first two PCA axes were accepted; variables that had similar factor scores were clustered together, with the highest scoring variable chosen to represent the cluster. Table S6 shows the selected variables (marked with a *) and the factor scores for the first two PCA axes for each climatic variable.

We then generated ENM for each species in *Camptoloma* and the genus as whole for each set of monthly values (i.e. 48 ENMs in total) using Hengl et al. (2009)'s approach as described above. These continuous predictions were run through a truncation threshold based on the DEC results to generate binary predictive models as described above (Fig. 2). A value of 0.7 was selected, with values ≥ 0.7 indicating habitat suitability and values < 0.7 indicating inhospitable habitat, according to present-day climatic requirements. We repeated this procedure with a "reduced present-day" dataset, which contained the recorded presence-only information and the monthly values for only two variables: mean temperature and precipitation (Appendix 8b). This was done to explore the ability of ENMs based on a reduced subset of data to generate geographic projections similar to those from the full "present-day" dataset. The palaeoclimate data used in the hindcasted ENM projections below include only climatic values for mean temperature and precipitation. If no differences are found between the habitat suitability models for the "full" versus "reduced" present-day projections, we can assume that the "reduced" ENMs are appropriate to perform the hindcasting of past distributions over time as the variables mean temperature and precipitation hold enough information to inform the projections. The same truncation threshold value as with the full dataset was used to transform continuous predictions into binary data for these "reduced" ENMs.

Hindcasted ENMs to generate past climate projections: We used palaeoclimate data composed of monthly values for two variables, mean temperature and precipitation, for three geological periods. These were obtained from global Hadley Centre-coupled ocean-atmosphere general circulation models (HadCM3L) that incorporate the effect of changes in atmospheric CO2 and represent major climate (warming or cooling) events worldwide (Beerling et al. 2009; Bradshaw et al. 2012; Meseguer et al. 2015): a) a 400 ppm CO_2 simulation with representation of Late Miocene palaeogeography representing the early warm period of the Miocene (Mid Miocene Climatic Optimum (MMCO, 15 Ma); b) a 280-ppm CO_2 Late Miocene simulation representing the cold and dry conditions prevalent after the Late Miocene Climate Cooling event (LMC, 11.6-5 Ma); c) a 560-ppm CO_2 Pliocene simulation representing the conditions at the Mid-Pliocene Warming event (MPWE, 3.6-2.6 Ma). We also included a simulation of the Preindustrial World with 280 ppm CO_2 to provide a baseline HadCM3L climate before the industrial revolution (Beerling et al. 2012). All data variables were normalised to lie between the closed boundaries [0-1] using the formula above. The present-day “reduced” ENMs, generated by month for each species and for the entire genus,

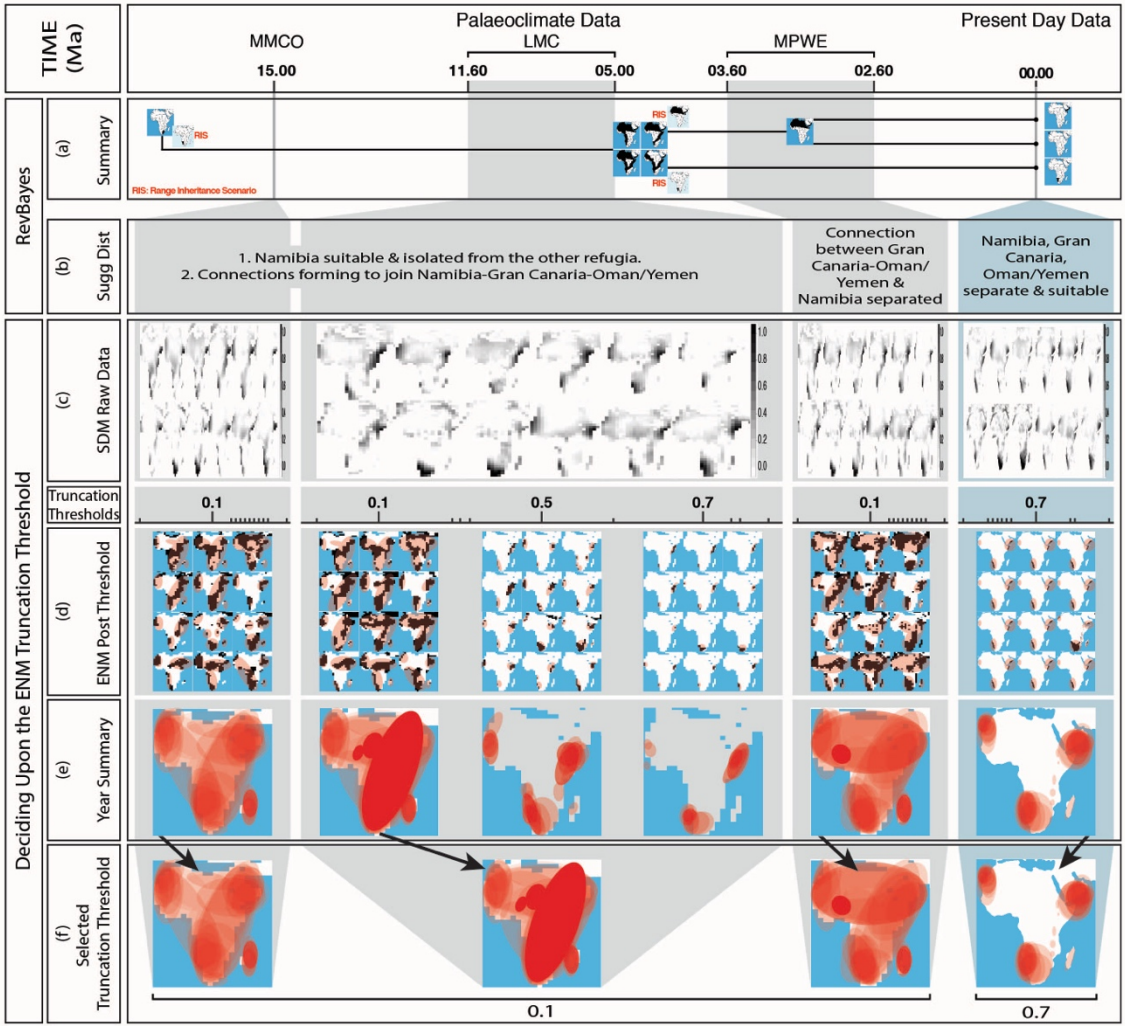


Figure 2. Phylogenetic-biogeographic approach to select the truncation threshold value for the ENM projections: The upper row shows the relevant climate and palaeoclimate layers used in the analyses. (a) MCC tree obtained from the BEAST nested-dating analysis, representing the stem- and crown-ancestors of *Camptoloma* (tree not to scale). Maps at tips show the current distribution of each species; maps at nodes represent the ancestral distributions inferred in the biogeographic analysis (RevBayes-DEC, Fig. 4); for some nodes, there could be several possibilities: "range inheritance scenarios", RIS). (b) "Sugg Dist" describes the potential distribution of ancestors during time periods along internal branches (i.e. between nodes). (c) "raw" monthly ENM models before applying the truncation threshold. Models are run over different truncation thresholds values, ranging from 0.1 to 0.9, in increments of 0.05 (scale). (d) "post" monthly ENMs after applying the truncation threshold; only those for 0.1, 0.5 and 0.7 are shown. Cells with values that fall below the truncation threshold are coloured in white ("uninhabited"); those equal to or greater than the threshold are painted in black ("inhabited"). The red circles represent clusters of connected suitable cells (suitable cells are considered connected if the distance between them is ≤ two unsuitable cells). (e): The 12-month projections with clusters are overlaid on top of each other to create a mean "year summary". (f) Selecting the truncation threshold: for hindcasted ENMs, we select the value that generates mean year summaries that best match the inferred ancestral distributions in the biogeographic analysis, this was 0.1 in our study. For present-day (preindustrial) and future ENMs, we select truncation threshold values that generate mean year summaries that are similar but not identical to the current distributions (0.7 in our analysis).

were hindcasted upon the three layers of palaeoclimate data. ENM palaeoclimate models were run through a truncation threshold to generate binary predictions as described in Figure 2. A value ≥ 0.1 was considered as indicating habitat suitability, with values < 0.1 indicating unsuitable habitat. The considerably lower value compared to present-day ENMs might be explained by the coarser spatial resolution of the palaeoclimate data (grid cells of 2.50°× 3.75°, around 100 x 100 km), and the uncertainty associated to projecting the climate model so far into the past.

Forecasted ENMs to generate future climate projections: We obtained from WorldClim 1.4 (www.worldclim.org/cmip5v1) four future climatic projections (downscaled and calibrated by WorldClim 1.4 i.e. bias corrected) from global climate models for greenhouse-gas scenarios (measurement representative concentration pathways; RCP=26, 45, 60, 80) for the year intervals 2041-2060 and 2070-2060. This data was first published in the Fifth Assessment IPCC report ($IPPC_{5,}$ www.ipcc.ch/assessment-report/ar5). The present-day "reduced" ENMs (only mean temperature and precipitation) were used to generate forecast ENM projections for each *Camptoloma* species and the genus as a whole. Predictions were run through the truncation threshold chosen for the present-day ENMs: grid cells with values ≥ 0.7 were considered climatically suitable for species; those < 0.7 were considered as inhospitable.

RESULTS

Phylogenetic and Biogeographic Inference

Phylogenetic analyses: Figure S2 depicts the consensus tree obtained in the MrBayes analysis (tree in TreeBase XXX). Figure 3 shows the corresponding BEAST tree (tree in TreeBase XXX). Both analyses recovered *Camptoloma* as monophyletic (PP=1), sister to tribes *Teedieae* and *Buddlejeae*; South African *Phygelius capensis* occupies a basal position within Teedieae. The main conflict between the two trees is the relationship among the three *Camptoloma* species*:* MrBayes recovers *C. canariense* as sister to *C. rotundifolium* and *C. lyperiiflorum* (Fig. S2), while in BEAST (Fig. 3), *C. rotundifolium* appears as sister to the clade *C. canariense*-*C. lyperiiflorum*.

Divergence time dating: The "all specimens" (Fig. 3, tree in TreeBase XXX) and the nested-dating MCC tree (shown in Fig. 4, tree in TreeBase XXX) in BEAST returned overlapping credibility intervals and similar mean ages for major lineages and the crown age of *Camptoloma* species. We follow here the nested-dating MCC tree since it accounts for biases introduced by the mixed species-population sampling. Divergence to *Camptoloma* from sister tribes Teedieae and

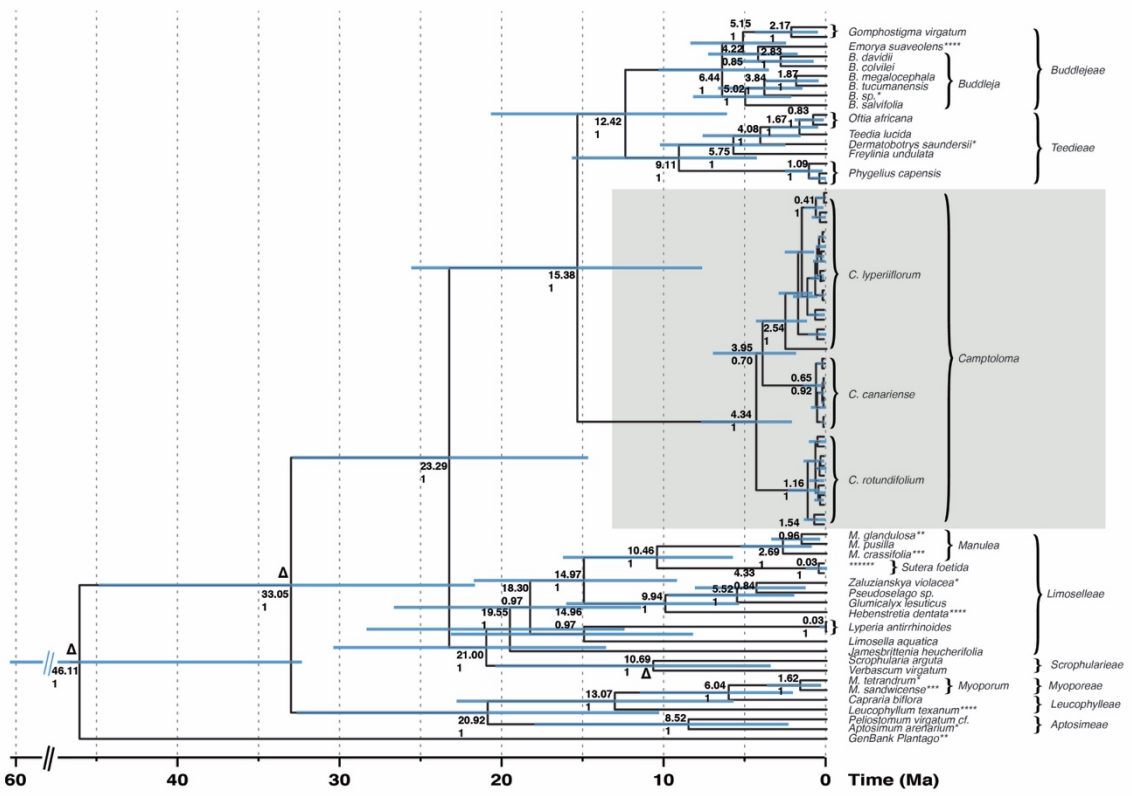


Figure 3. BEAST maximum clade credibility (MCC) tree based on the "all specimen dataset", showing 95% High Posterior Density (HPD) credibility intervals in blue (Δ indicates nodes constrained by secondary time calibrations). Numbers above branches are mean age estimates; numbers below represent clade posterior probability values (PP). Symbol * indicates the number of markers that failed sequencing for that particular taxon.

Buddlejeae (stem-node) was estimated in the Mid Miocene (21.1 Ma, 95% HPD: 11.52-34.05). The crown-age was estimated in the Miocene-Pliocene boundary, at 4.26 Ma (95% HPD: 1.29-9.09 Ma), while the split between *C. canariense* and *C. lyperiiflorum* was estimated at 3.42 Ma (1.05-7.70). The start of population-divergence (species crown-node) was dated in the Pleistocene for *Camptoloma rotundifolium* (1.85 Ma; 0.61-4.24) and close to the Plio-Pleistocene boundary for *C. lyperiiflorum* (3.09 Ma; 1.28-6.11), while considerably younger in *C canariense* (0.76 Ma; 0.23-1.69).

Biogeographic and phylogeographic analyses: The RevBayes-DEC analysis (Fig. 4, tree in TreeBase XXX) reconstructed the most-recent common ancestor (MRCA) of *Camptoloma* and sister tribes *Teedieae* and *Buddlejeae* as occurring in eastern South Africa. This was followed by migration to the west in *Teedieae*, and to the west and north in *Buddlejeae*; the current cosmopolitan distribution of *Buddleja* and *Emorya* is explained by dispersal from southern African ancestors. The stem-ancestor of *Camptoloma* was inferred as occurring in Namibia (western South Africa, Fig. 4), implying a dispersal event from the east. Sister tribes Limoselleae

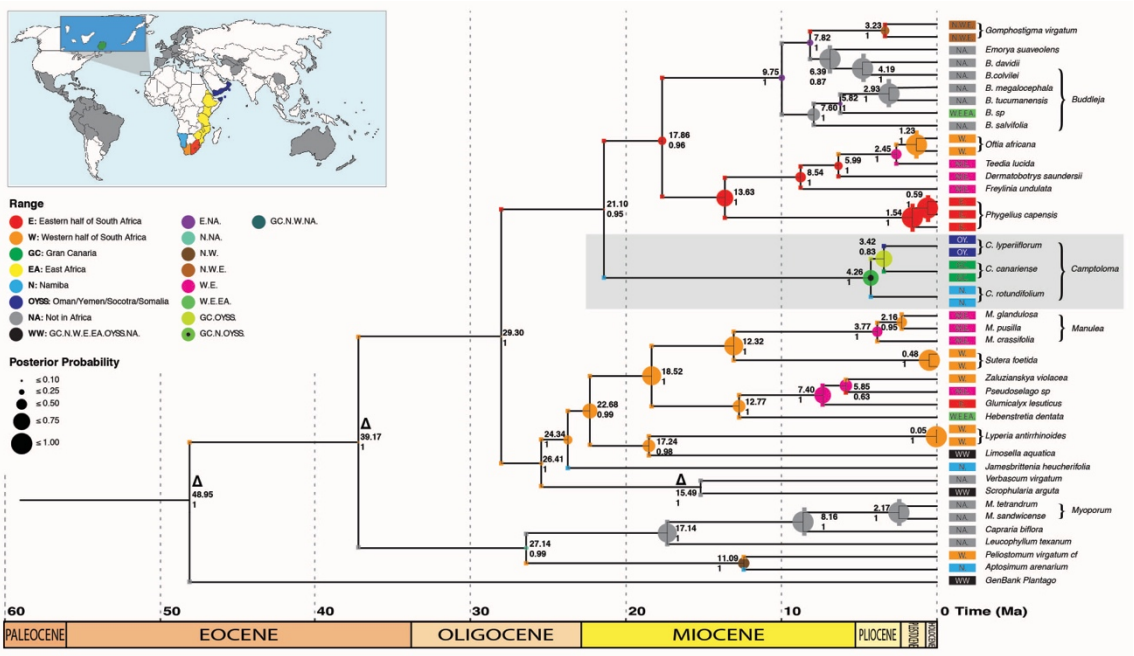


Figure 4. Bayesian inference reconstruction of the biogeographic history of *Camptoloma* and relatives in Scrophulariaceae, using the Dispersal-Extinction-Cladogenesis (DEC) model implemented in RevBayes. The phylogeny is the MCC tree from the nested dating analysis; numbers above branches indicate mean ages; those below branches represent posterior probability values (PP), with Δ indicating nodes constrained by secondary time calibrations. Coloured rectangles close to taxon names indicate the current distribution of each taxon (see inset map and legend for colour codes; widespread distributions are represented by combining single-area colours, e.g. the pink W.E. is represented as red and orange in the map). Range inheritance scenarios are presented at each node as squares: the size of the circles is proportional to the marginal posterior probability of the inferred ancestral range (see inset legend).

and Scrophularieae are also of western South African origin. However, inferences for many of these backbone nodes received very low marginal posterior probabilities (< 0.01), a result of the limited taxon sampling, long internal branches and the cosmopolitan distribution of Scrophulariaceae. Crown-node *Camptoloma* was reconstructed as already occupying the three regions where the three extant species occur (> 0.75), implying a dispersal event northwest (Gran Canaria) and eastward (Oman/Yemen/Socotra/Somalia). The current disjunct distribution of the genus is explained by a first vicariance event between southern and northern ancestors, followed by vicariance between west and east across the Sahara.

Compared to BASTA, the DTA analysis returned ancestral location inferences with higher posterior probability values (Fig. 5). Results for *C. canariense* were very similar between the two methods, depicting geographically structured patterns, with adjacent populations clustered together. For *C. lyperiiflorum*, and especially for *C. rotundifolium*, BASTA favoured ancestral locations and source areas of migration events in areas underrepresented or occupying a

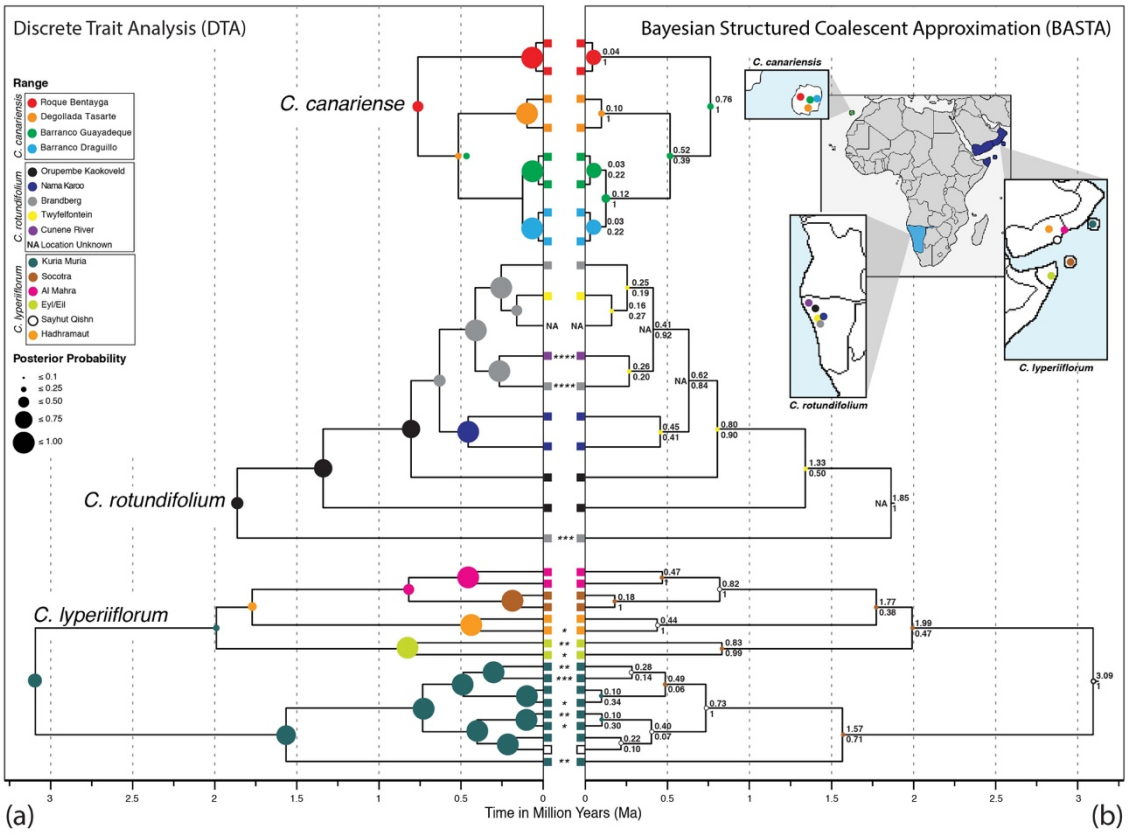


Figure 5. Bayesian inference reconstruction of the phylogeographic history of the three *Camptoloma* species, obtained with: A) Discrete Trait analysis (DTA) and B) Bayesian Structured Coalescent Approximation (BASTA). Numbers above branches indicate mean ages; those below branches represent clade posterior probability values (PP). The tree topology and branch lengths were constrained to follow the nested-dating MCC trees. Maps on the right show location of current populations, with colour codes and legends indicated in the left inset legend. Coloured squares at tips indicate the distribution of each individual/sequences. Circles at nodes depict the inferred ancestral range; the size is proportional to the marginal posterior probability (see left inset legend). Symbol * indicates the number of markers that failed sequencing for that particular taxon.

derived position in the tree (e.g. Twyfelfontein in "yellow" for *C. rotundifolium* or Yemen (Sayhut Quishn) in white for *C. lyperiiflorum;* Fig. 5). DTA, however, showed much more conserved geographical patterns, implying a lower frequency of migration events (Fig. 5), and exhibited larger PP nodal values. Dated nested population, BASTA and DEC tree files can be found in in TreeBase XXX.

Ecological Niche Models and Geographical Projections

Ecological Principal Component Analysis: The e-PCA analysis (Fig. 6) showed a clear segregation among the three species of *Camptoloma* in the environmental space (Table S7 lists the factor scores for each variable). Based on the length of (PC1, PC2) to the origin, the most explanatory variables differed among species: yearly water vapour pressure (VAPR yearly) and mean temperature of the driest

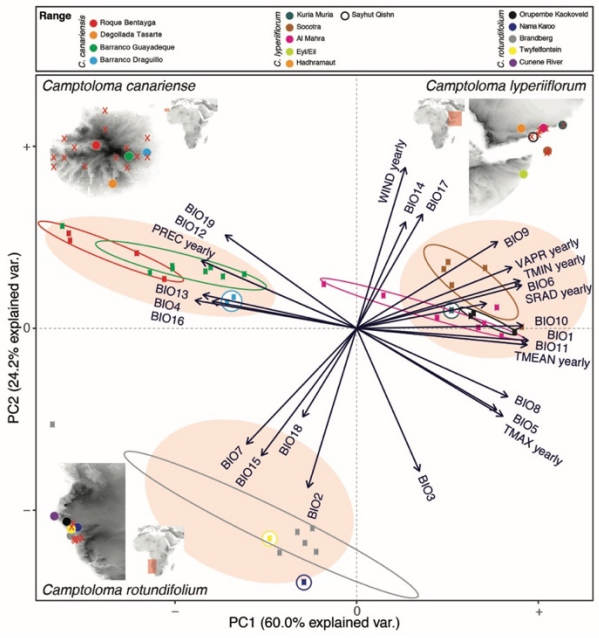


Figure 6. Environmental Principal Component Analyses (ePCA), showing segregation of the three *Camptoloma* species over the climatic space along the first two axes, PC1 and PC2 explaining 84% of variance. Codes for climatic variables are given in Table S5; factor scores are listed in Table S7. Sampled individuals' populations are indicated by the colour of their square and by coloured circles on the maps, with colours represented in the inset legend, following the operational areas used in the phylogeographic analyses (Fig. 5). Pink, solid circles group the square points together by *Camptoloma* species and coloured circles group square points together by population.

quarter (BIO9) for *C. lyperiiflorum*; yearly precipitation (PREC yearly) and quarter (BIO16) for *C. canariense*; and the relationship between precipitation and temperature (precipitation seasonality, BIO15) and the mean diurnal range (mean of monthly (max temperature – min temperature), BIO2) for *C. rotundifolium*. Camptoloma *canariense* and *C. lyperiiflorum* were segregated along the first e-PCA axis (explaining 60% of the variation in the environmental space), with *C. canariense* preferring a wetter and colder niche than *C. lyperiiflorum*. *Camptoloma canariense* appears separated from *C. rotundifolium* along the second e-PCA axis, which explains 24.2% of the environmental variation, with *C. canariense* preferring a colder niche than *C. rotundifolium*. No clear pattern of segregation or clustering was found at the population level (Fig 6, insets), especially within *C. rotundifolium*.

***Present-day ENMs**:* Table S6 lists the factor scores for each variables per month, * marks the variables retained after the PCA performed on the "full present-day" dataset. Figure S3 and Figures S4 (a, b) show the raw ENM projections for the "full" and "reduced" models before using the truncation threshold for each species and the genus. Figure 7 shows the corresponding "summary" ENM post-truncation projections for the genus. All projections agree in showing a much larger extension of suitable areas compared to the current distribution of each species (Fig. 1). They also predict the presence of suitable habitat for all three species across Africa and Madagascar for at least one month of the year, including habitat suitability across the Sahara Desert for *C. rotundifolium* and *C. lyperiiflorum* (in the reduced model only, Figs. S4 a, b). Yet, suitable habitat throughout the year was only predicted in those areas where species are found today (Figs. 1, 7). Some differences were found among species between the full and reduced datasets (Figs. 7, S3, S4a, b). For example, ENMs based on the reduced

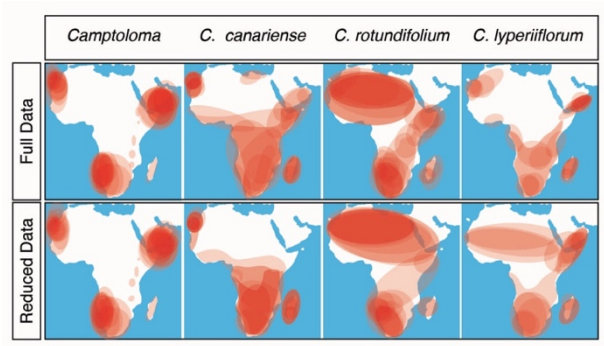


Figure 7. Present-day ENMs: Geographical projections of the "full" and "reduced" ENMs for genus *Camptoloma* and each individual species, based on present-day climatic data. These are post ENM projections after applying a truncation threshold of 0.7. Table S5 lists the climatic variables employed after removing correlated variables through an ePCA step. Red indicates habitat suitability and white habitat unsuitability. Red circles show clusters of suitable habitat after overlaying the 12 monthly projections to generate a mean year summary (see text for further explanation).

dataset predict some habitat suitability across the Sahara Desert for *C. lyperiiflorum* that is not recovered in the full model; conversely, the full model shows suitable habitat for *C. canariense* in southern Arabia not captured by the reduced model

(Fig. 7). Yet, in general there was a good agreement between the two sets of projections and excellent agreement between the two projection for *Camptoloma* as a whole. Because of the "excellent agreement", this lends support to our methodological assumption that the reduced dataset contains enough information to describe habitat suitability in *Camptoloma*, and hence the hindcasted-reduced models can also be used to make reliable predictions. Figure S5 shows the ENM models using the full dataset but at the population level (within-species), results were uninformative.

Hindcasted ENMs: Figure 8 shows the results of the summary reduced ENMs for *Camptoloma* hindcasted over the palaeoclimate layers. Figures 4 (c, d) show the corresponding raw ENM predictions for *Camptoloma* before applying the truncation threshold. The first observation to notice is that the hindcasted ENMs show connections of suitable habitat among the three present-day disjunct distributions. These connections, however, can only be observed in some of the monthly projections and change across the three geological periods (Fig. 8); if ENMs are "averaged" across months to produce a yearly summary, these connections are not observed to be as strong. A second observation is that the ENM projections for the pre-industrial world (0 Ma) showed habitat suitability restricted to the present-day disjunct distribution of the species, except that the western projections cover also the western Sahara/Moroccan coast.

Forecasting ENMs: Figure 9 shows the summary reduced ENM projections for the four global warming scenarios (Figure S6 shows the corresponding results before the values were run through the truncation threshold). When

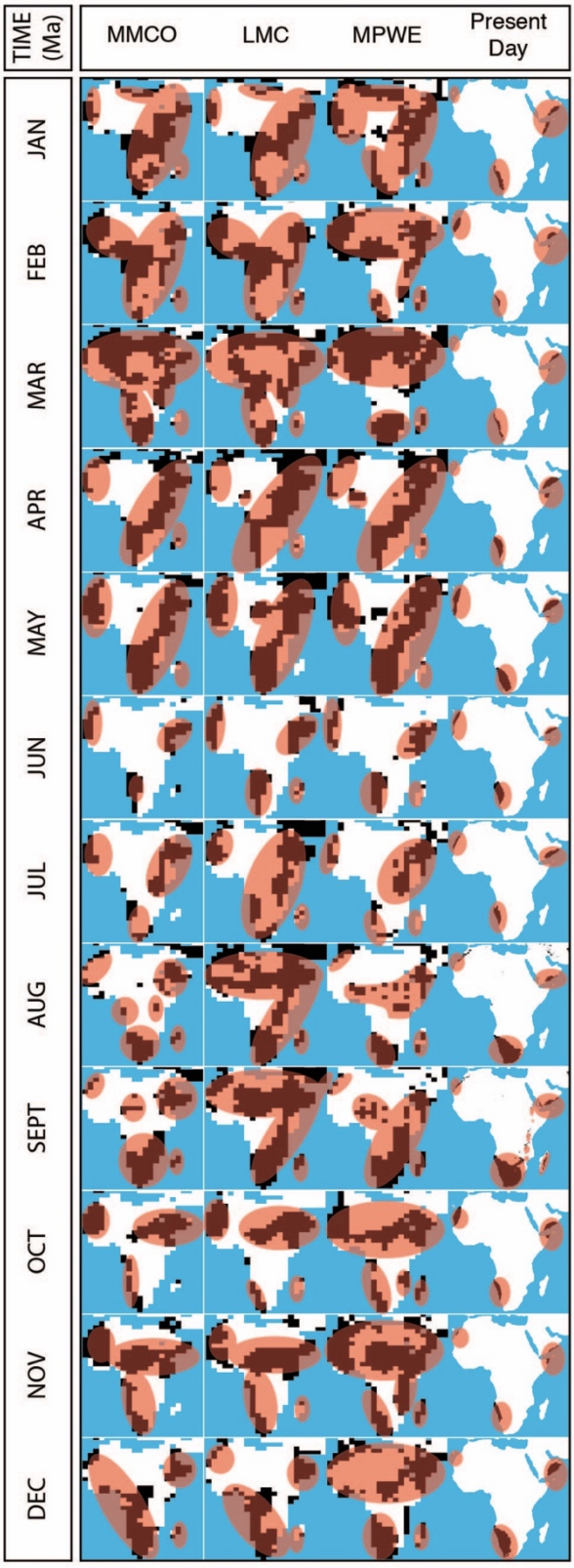


Figure 8. Hindcasted ENMs: Geographical projections of the reduced ENMs for *Camptoloma* and each individual species, based on the palaeoclimate data for four periods: i) Mid Miocene Climate Optimum (MMCO, 15 Ma); ii) Late Miocene Climate Cooling, (LMC, 11.6-5 Ma); iii) Mid Pliocene Climate Warming (MPWE, 3.6-2.6 Ma), and iv) preindustrial present day (0 Ma). These are post ENM projections after applying a truncation threshold of 0.1 for palaeoclimate layers and 0.7 for the preindustrial present day. Black indicates suitable habitat and white, unsuitable habitat. The red circles represent clusters of connected suitable habitat in those time periods, after overlaying the 12 monthly projections to generate a mean year summary (i.e. when the distance between them is ≤ two unsuitable cells, Fig. 2).

comparing the predicted habitat suitability with the present-day projections (Fig. 7), no differences can be seen.

Combining ENM projections with phylogenetic biogeographic reconstructions: Figure 10 shows how the ancestral inferences from the RevBayes DEC analysis (Figs. 4, 10a) can be combined with the ENM hindcasted projections (Fig. 8) to produce an informative picture of the evolution of *Camptoloma* across geographical space and over geological time. The ENMs' "hindcasting summary" (Fig. 10b) show the monthly projections in Fig. 8 "overlaid" (instead of averaged) across the 12 months to produce a "yearly" summary, to better display the connections among suitable habitats over time. Comparison with the RevBayes-DEC ancestral range inferences (Fig. 10a) shows a good match, summarised in Fig. 10c. For example, DEC reconstructs the MRCA of *C. canariense* and *C. lyperiiflorum* (3.42 Ma) as occurring in Gran Canaria and Oman/Yemen/Socotra/Somalia, implying a widespread ancestor across northern Africa

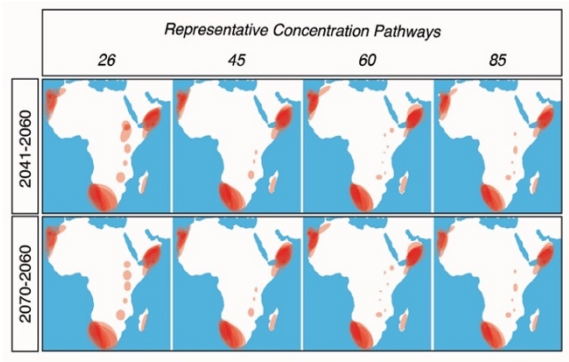


Figure 9. Forecasted ENMs: Geographical projections of the reduced ENM models for genus *Camptoloma* forecasted onto four global warming scenarios for year intervals 2041-2060 and 2061-2080. These are post ENM projections after applying a truncation threshold of 0.7. All other conventions as in Figure 8.

divided by vicariance (Figs. 4, 10a). This is supported by the hindcasted ENM projections for the Mid Pliocene warming event, which show full connections of habitat suitability across the Sahara for several monthly projections (e.g. February), and partial connections across other months (e.g. May) (Fig. 8), summarised in the MPWE of Fig. 10b. Similarly, RevBayes infers the MRCA of the three species (4.26 Ma) to cover a widespread distribution spanning southwest Africa, Oman-Yemen and Gran Canaria. This contrasts with the inferred geographically restricted stem-ancestor (21 Ma) in southwest Africa (Figs. 4, 10a), suggesting a dispersal event eastwards and north westwards over the 15 million year branch-length separating crown and stem *Camptoloma* from its stem-ancestor. The ENM projections for the Late Miocene period (LMC, 11.6-5 Ma) and the Mid Miocene Climatic Optimum (MMCO, 15 Ma), which lie within the transitional period between the two nodes (21 to 5 Ma), provide evidence for these transient biotic connections (Fig. 10b). The LMC projections show corridors of habitat suitability spanning diagonally from South Africa/Namibia to Oman/Yemen/Somalia for several months, whereas Gran Canaria is connected to Oman/Yemen/Socotra/Somalia in some projections (Fig. 8). A similar pattern can be found in the MMCO ENMs, except that connections between southern and eastern Africa appear more frequently in the monthly projections, whereas connections across the Sahara and northern Africa are seen only across a few months (Fig. 8). Interestingly, some ENM projections for this time slice show connections between southwest and northwest Africa.

DISCUSSION

Phylogenetic Position of *Camptoloma* and the Origin of the Rand Flora Disjunction

Scrophulariaceae is a large angiosperm family, which formerly included genera that are currently part of *Plantaginaceae* and has been the subject of intense taxonomic revision due to conflict between morphological and molecular characters (Kornhall et al. 2001; Oxelman et al. 2005). Genus *Camptoloma* was initially included in tribe Manuleae based on floral morphology (Hilliard 1994). In the first molecular phylogenetic analysis of this tribe, Kornhall et al. (2001) excluded *Camptoloma*

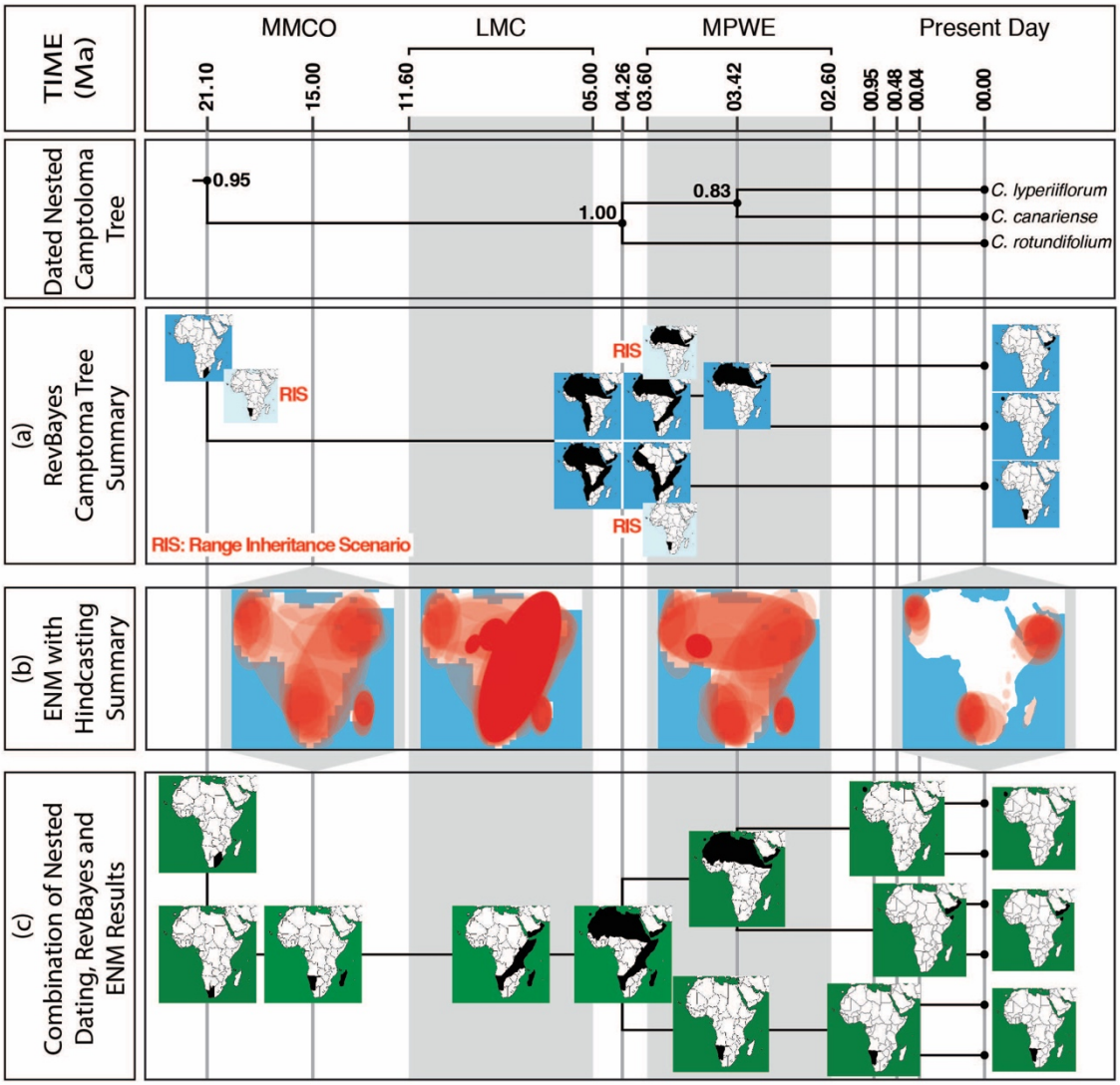


Figure 10. Combining information from ENMs and biogeographic inference to reconstruct the spatio-temporal evolution of *Camptoloma*: Shown at the top is the MCC tree from the BEAST nested-dating analysis, representing the stem- and crown-ancestors of *Camptoloma* (tree not to scale), with mean age estimates and PP values. (a) RevBayes-DEC reconstruction, depicting the species current distribution and inferred ancestral distributions at nodes, as in Figure 2a. (b) Post-truncation ENM projections, representing the mean year summaries for the three palaeoclimate layers and the present-day data (see Fig. 2 for more details on how these are constructed). (c) The information provided by the ENMs is used to fill the gaps in the biogeographic DEC analysis. For example, hindcasted projections for the MMCO and LCM periods –which are temporally bounded by the long branch (15 Ma) separating the stem- and crown-ancestors of *Camptoloma*, suggest the existence of climatic corridors with connected areas of habitat suitability between southwest Africa and northeast Africa. Similarly, ENM projections for the later MPWE period depict potential connections across northern Africa, in agreement with DEC inference (Fig. 2a), suggesting that the E-W disjunction was formed after the connection with southern Africa was interrupted. Notice that these biogeographic corridors can only be observed in some of the monthly geographic projections. Thus, averaging ENMs across months, as commonly done in ENM present-day projections, fails to show these connections; only when monthly projections are overlaid on top of each other as done here (Fig. 2b), the full pattern can be observed.

and synonymised Manuleae with tribe Limoselleae. Oxelman et al. (2005) supported this classification and placed *Camptoloma* as sister to the mainly southern African tribes *Teedieae* (10 species) and *Buddlejeae* (c. 100 species). The three form a clade with tribe *Scrophularieae* and an extended tribe *Limoselleae* (Oxelman et al. 2005). A recent phylogenomic study on Buddlejeae supports this arrangement (Chau et al. 2018).

Although some small differences can be seen at the level of genera, our combined nuclear-cpDNA phylogeny (Fig. 3) confirms these phylogenetic relationships, showing *Camptoloma* as sister to Teedieae and Buddlejeae, with *Phygellius capensis* basal to the other Teedieae genera (contrary to Pokorny et al. 2015). Among the three *Camptoloma* species, relationships are not conclusive, despite including ITS and new cpDNA markers and a larger intraspecific sampling. The BEAST MCC tree (Fig. 3) supports identical relationships as Kornhall et al. (2001) and Oxelman et al. (2005), showing *Camptoloma rotundifolium* as sister to the clade *C. canariense*-*C. lyperiiflorum*. The MrBayes tree (Fig. S2) shows a similar topology to Pokorny et al. (2015), with the Canarian endemic sister to the other two species. Differences in tree-growth and branch length priors might explain this difference. We favour here the BEAST topology (used in the biogeographic analysis, Fig. 4) because: a) it agrees with Kornhall et al. (2001) and Oxelman et al. (2005) cpDNA phylogenies (based on a different set of markers), and b) is congruent with preliminary results from a phylogenomic study on Scrophulariaceae, targeting hundreds of low-copy nuclear markers (Villaverde et al. 2019). This topology was used in the biogeographic analysis (Fig. 4).

Our spatiotemporal scenario reconstructs the stem-ancestor of *Camptoloma* as distributed in southwestern Africa in the Early-Mid Miocene (c. 21 Ma, Fig. 4). During this period, changes in global atmospheric circulation profoundly transformed the climate of the continent (Griffin 2002), which could have limited biotic connections between northern and southern Africa (Sanmartín et al. 2010). The northeastward drift of the continent, the closing of the Tethys Seaway by the Arabian Plate, and the gradual uplift of Eastern Africa introduced a drier and more arid climate (Sepulchre et al. 2006; Trauth et al. 2009; Pokorny et al. 2015). Desertification started in the southwest (17-16 Ma), with the formation of the current Namib Desert, and advanced northeastwards, with the expansion of savannahs in eastern Africa (8-9 Ma) and the appearance of the large Sahara Desert in North Africa (7 Ma; Senut et al. 2009).

The Late Miocene period (11.5-6.5 Ma; Beerling et al. 2009) was characterised globally by colder and wetter conditions (Zachos et al. 2001). Mairal et al. (2017a) predicted that this period would have been used by RF lineages to disperse northward and westward. Our MMCO

and LMC hindcasted ENM models support the existence of climatic corridors of suitable habitat connecting northern and southern Africa along Eastern Africa (Fig. 10b). Though there are some projections showing connections across western Africa (Fig. 8), most monthly projections support an eastward corridor. This route can also be glimpsed in the present (Fig. 7) and future (Fig. 9) ENM projections, as disconnected patches of suitable habitat. This eastward migration route agrees with other studies of Afrotemperate groups (Galley et al. 2007; Popp et al. 2008; Roquet et al. 2009; Barres et al. 2013; Meseguer et al. 2015; Mairal et al. 2015a). García-Aloy et al. (2017) supports a similar scenario for the African genus *Monsonia* (Geranieaceae), with an origin in southwestern Africa in the Early Miocene (c. 21 Ma), followed by northeastward dispersal (4-6 Ma), coincident with a general cooling trend and the Early Pliocene uplift of the Eastern Arc Mountains (c. 5 Ma, Sepulchre et al. 2006).

Although in northern Africa the first signs of aridification appeared around 7 million years ago, this was initially in small patches. The Early Pliocene was a relatively humid and colder period, but aridification intensified during the global climate-warming event of the Mid Pliocene (3.5 Ma; Zachos et al. 2001). The MPWE was a consequence of global (e.g. the closing of the Panama Isthmus) and regional tectonic events, e.g. the uplift of the East Africa Rift System, which created a rain shadow between central-west Africa and the drier east African plateau (Sepulchre et al. 2006). These events likely interrupted the biotic connections between southern and northeastern Africa, which are observed in the LMC projections but disappear in the hindcasted ENMs for the MPWE (Figs. 8, 10b). On the other hand, this later palaeoclimate layer shows connections between northwest and northeast Africa, which might account for a subsequent expansion of the ancestor of *C. lyperiiflorum* and *C. canariense* (3.42 Ma, Fig. 4) across North Africa. There is, however, a large uncertainty in these age estimates, with long credibility intervals going back to the Late Miocene-Early Pliocene (Figs. 3, 4); these migration events could therefore have taken place earlier, when climates were more humid, or even later. Connections across the Sahara disappear in the pre-industrial ENM projections (Figs. 8, 10b), in line with the tendency towards a more xeric flora in northern Africa from the Pleistocene onwards (De Menocal 2004; Mairal et al. 2017).

The scenario described above supports the "climatic refugia" hypothesis, in which the current disjunct distribution of *Camptoloma* across Africa is the result of fragmentation and extinction/population bottlenecking events associated to increasing aridification (Sanmartín et al. 2010; Mairal et al. 2015a, 2017a). An alternative explanation is that these continental disjunctions are the result of long distance dispersal (LDD) events. LDD has been

argued for some similarly wide-ranged disjunctions in Africa (e.g. Pirie et al. 2018) and worldwide (Popp et al. 2011). Yet, the deep temporal divergences among the three *Camptoloma* species (Early-Mid Pliocene) and their rarity status (small populations) stands in contrast with other African disjunct genera, where dispersal over long distances across Africa has been followed by niche release and rapid diversification (Garcia-Aloy et al. 2017; Pirie et al. 2018). An alternative scenario is the one argued for another widely disjunct, rare RF genus, *Canarina* (Fig. S1; Mairal et al. 2015a). Instead of LDD or a past continuous distribution across Africa, short-distance dispersal and stepping-stone colonisation, taking advantage of continental changes in topography and climate (Pirie et al. 2018), and followed by extinction, could explain the wide-continental RF disjunctions. Indeed, *Camptoloma* shows a similar pattern to *Canarina*, with the Canarian endemic exhibiting a younger age for the start of population divergence than the two African endemics (Figs. 3, 5; Mairal et al. 2015a). Similar results have been reported for *Euphorbia balsamifera* (Euphorbiaceae, Villaverde et al. 2018) and *Plocama* (Rubiaceae, unpublished results). One explanation is that the Macaronesian component of the Rand Flora is of recent origin: the result of a dispersal event by a northwestern African population that went later extinct, leaving a spatial "gap" between the east and western disjunct taxa (Mairal et al. 2015a; Villaverde et al. 2018).

At species level, the phylogeographic analyses (Fig. 5) did not provide any conclusive evidence. While phylogeographic structure was very conserved within species according to DTA, BASTA indicated frequent dispersal events from rarely represented populations, especially for *Camptoloma rotundifolium*. The latter could be a bias produced by the low overall migration rate and the highly unbalanced tree; under these conditions, BASTA can recover anomalous bimodal distributions (Cornuault and Sanmartín 2019). The higher PP support for DTA inferred ranges compared to BASTA is probably a bias resulting from assuming a mutation-only process (De Maio et al. 2015). Despite these differences, both methods agree on some interesting events. An origin of *C. canariense* populations from central (BASTA) or western (DTA) areas of Gran Canaria agrees with the long-persistence of plant lineages in these areas, which have a recent history of geological stability (Emerson 2003; Mairal et al. 2015b). In *Camptoloma lyperiiflorum*, long-term persistence is supported in island systems (Al-Hallaniyah in DTA, Socotra in BASTA), which would have maintained a milder climate than the surrounding arid regions (Domina et al. 2012).

Nowadays, *Camptoloma* populations occupy climatic microrefugia, such as areas of high humidity and low seasonality in the

Socotra Archipelago (*C. lyperiiflorum*, Domina et al. 2012), shaded vertical crevices in Gran Canaria (*C. canariense*), or rocky crevices in the Namibian coastal Brandberg mountain range (*C. rotundifolium*, Craven and Craven 2000). The three species appear segregated in the environmental space (Fig. 6) and have been isolated for several million years (Fig. 3), which agrees with the hypothesis of climatic relicts. Macroecological studies suggest that thermal tolerances to maximum temperatures are more conserved than to minimum temperatures over long-term evolutionary history (Araújo et al. 2013; Saupe et al. 2017; but see Pellissier et al. 2013). If rates of niche evolution are lower for maximum temperatures than for other climatic variables, developing adaptations to increasingly warmer climates might be problematic for plant lineages. This has implications for the origin of the disjunct RF taxa and for their survival under scenarios of global climate warming. Yet, our forecast ENM projections for *Camptoloma* (Fig. 9) do not show major changes in the current distributions, which might indicate that current refugia might remain stable under future climate change.

Combining Ecological Niche Modelling with Phylogenetic Biogeography to Reconstruct the Role of Ancient Climate Change in Depauperate Lineages

Truncation thresholds are used to transform the continuous predictions of ENMs into presence/absence data in order to establish a species' geographic range –a landscape of habitat suitability– based on occurrence records (Diniz-Filho et al. 2010). Typically, truncation thresholds are selected using model fit tests, such as the Kappa statistic (Monserud and Leemans 1992; Fielding and Bell 1997), the Area Under the Curve (AUC) or the Receiver Operator Characteristic (ROC) curves. Recent methods such as the true skill statistic (TSS, Allouche et al. 2006) or the leave-one-out test (Pearson et al. 2007) were developed to compensate for potential shortcomings in the Kappa (Allouche et al. 2006) and the AUC methods (Lobo et al. 2008). All these approaches rely on using a subset of the original data, the "test or training dataset", to evaluate model fit through examination of true and false predictions of presences and absences of a species (Liu et al. 2005; Jiménez-Valverde and Lobo, 2007). When the original data is complete and unbiased, these methods of model fit choice are reliable.

On the other hand, it is often the case when working with real data that presence data is incomplete, biased towards certain regions or countries due to sampling effort, or that, for certain organisms, it might be difficult to distinguish between native and invasive ranges. Data incompleteness is a special concern in RF lineages. Compared to other geographic regions, Africa has lagged behind in phylogenetic and taxonomic studies, with new records of described or new taxa not infrequent

(Couvreur et al. 2014; Villaverde et al. 2018). In addition, the RF pattern involves large disjunctions across difficult-to-access or politically unstable regions (e.g. Namibia, Yemen or Somalia), which has resulted in a strong bias in occurrence representation in online databases such as GBIF. Methods such as ENFA or Mahalanobis Distance can be used when only presence data is available (Basille et al. 2008; Clark et al. 1998; Meseguer et al. 2015). For methods that require presence/absence data, the treatment of pseudoabsences is an important issue that can mislead ENM projections (Chefaoui and Lobo 2008). Dealing with pseudoabsences is particularly relevant in *Camptoloma* because of the very wide disjunct distribution and paucity of occurrence records. In this study, we employ an approach, first developed by Engler et al. (2004) and later expanded by Hengl et al. (2009), which combines the best characteristics of presence/absence models and habitat suitability distances into one framework. In particular, habitat suitability distances obtained from ENFA are used to weight the sampling when creating pseudoabsences, which are then used in creating a ridge regression GLM (Hengl et al. 2009).

To select the truncation threshold for transforming the ENM continuous predictions into a binary model of habitat suitability/unsuitability, we combined information from a further source of evidence: the phylogenetic-biogeographic inferences returned by the RevBayes-DEC analysis (Figs. 2, 4). One advantage of this approach is that, unlike in model fitting tests such as TSS, the input data used in DEC (i.e., the molecular time tree and user-defined areas) are independent from the original data used for generating the ENMs (i.e., the occurrence records and climatic data). Another advantage is that using biogeographic information allows us to have different thresholds for the palaeoclimate (0.1) and present-day (0.7) projections. This is important because the coarser resolution of the palaeoclimatic data implies that that we could not have captured any pattern if we would have applied the stricter truncation threshold used for the present-day projections (Fig. 2). This emphasizes the importance of selecting an appropriate truncation threshold when using presence/absence ENM models (Nenzén and Araújo 2011).

If biogeographic inferences can be used to calibrate ENM models as described above, ENMs might likewise be employed to inform and complement biogeographic inference, especially when there is incomplete or sparse phylogenetic data. DEC biogeographic inferences are reliant on the accuracy of the user-defined operational areas to describe the species' current distribution. If these are biased (e.g. incomplete taxon or geographic sampling, invasive species ranges), it could mislead biogeographic reconstructions. High extinction rates, through removal of certain nodes and lineages, especially if area-dependent, can also

lead to erroneous biogeographic inferences (Meseguer et al. 2015; Sanmartín and Meseguer 2016). Phylogenies with a broom and handle shape have been associated with high background extinction rates or mass extinction events (Antonelli and Sanmartín 2011; Sanmartín and Meseguer 2016). *Camptoloma* exhibits a very long branch of length c. 15 million years between the stem divergence from its sister-lineage (tribes Teedieae and Buddlejeae) and the start of crown-group diversification (Figs. 3, 4). Though parametric biogeographic methods like DEC account for branch length information, the absence of speciation nodes or diversification events introduces statistical uncertainty in the inference of ancestral ranges (as evidenced by the low PP values).

ENM models provide a spatial geographic framework, which is absent from parametric biogeographic analysis. Some authors have advocated integrating evidence from ENM models to calibrate biogeographic/phylogeographic inference, either to include additional areas of suitable habitat where a lineage has gone extinct or has yet to be found, or to evaluate dispersal or migration models by postulating climatic corridors connecting areas across unsuitable habitats (Richards et al. 2007; Smith and Donoghue 2010; Meseguer et al. 2015). Incorporating information from the fossil record is another source of evidence, especially in cases of high extinction rates. However, many of these macroevolutionary models require large or medium-sized phylogenies (> 20 tips, Sanmartín and Meseguer 2016). This is especially problematic for RF lineages, which often include less than 10 tips and lack a fossil record (Pokorny et al. 2015). Here, we propose a different approach in which the two methods, ENMs and biogeographic analysis, are used as independent but complementary sources of evidence. We used ENM hindcasted projections to provide additional evidence for nodes receiving low support in the RevBayes-DEC ancestral area inference, for example, the origin of stem-*Camptoloma* in southwestern Africa (Figs. 4, 8). ENMs also helped to complement the biogeographic inference along the long internodes connecting the species and the stem-lineage, providing support for the existence of ancestral climatic corridors between now disjunct regions, and, hence, for the climatic-fragmentation/aridification-extinction hypothesis (Fig. 10c).

There are several potential shortcomings with our approach. One is that ENM models assume conservation of climatic niches, or in other words, that the ancestral niche is the sum of the extant species individual niches. This strong assumption could be alleviated by including fossil information (Meseguer et al. 2015; 2018), but this is absent or unknown in *Camptoloma*. The small size of the phylogeny makes also unviable to use comparative phylogenetic methods of ancestral niche evolution (Meseguer et al. 2018). There

is still debate about conservation of ecological niches over evolutionary time scales (Peterson 2006), especially under scenarios of rapid climate change. We do not make here any inferences about conservation or evolution of niches, but use ENMs only as a mean to complement the DEC inferences. Also, ENMs can only provide estimates of the fundamental ecological niche, that is, the set of climatic conditions under which the species can maintain viable populations (Peterson 2006). The realised niche could be different, a result of dispersal barriers and interactions with other species (Saupe et al. 2017). In this sense, biogeographic inferences with RevBayes-DEC might help discriminate between the fundamental and the realised niche (Saupe et al. 2017); for example, ENMs predict the presence of *Camptoloma* ancestors in Madagascar (Fig. 8), where no species of the genus occurs today (Fig. 1), suggesting that other factors such as dispersal barriers or interactions with other species might have prevented the genus to occupy the full extent of thee potentially suitable geographical space (Saupe et al. 2017). Finally, the use of monthly ENM predictions to generate a yearlong summary might seem unrealistic. While considering monthly climatic conditions makes sense in angiosperms, whose life cycle and phenology (e.g., seed dispersal, pollination) is strongly associated to changing and seasonal climatic conditions (Du et al. 2015), doing this for palaeoclimate scenarios going back millions of years appears risky. In our study, biogeographic corridors can only be observed in some of the monthly ENM projections (Figure 2b). Averaging ENMs across months to generate a "yearly" mean value –as it is commonly done in ENM studies (Meseguer et al. 2015; 2018; Mairal et al. 2017)– results in a loss of signal. Only when monthly projections are overlaid on top of each other (Figs. 2b, 10c), is the full pattern recovered. Araújo and New (2007) reached a similar conclusion when they argued that the combination of forecasts, each containing some independent information, as in "ensemble forecasting", would yield lower mean error than any of the constituent individual forecasts. One potential explanation for this is that averaging across projections without accounting for error increases uncertainty in the data, destroying or flattening out any remaining signal. In our case, there are multiple sources of uncertainty, including the use of a reduced climatic dataset, the coarse resolution of the palaeoclimate layers, and the incompleteness of the data itself. The fact that our ENM hindcasted summary projections agree with the biogeographic analysis adds reliability to our inferred evolutionary scenario (Fig. 10c).

CONCLUSIONS

In this study, we showed that species-poor lineages in which there is a dearth of data, such as Rand Flora *Camptoloma*, the combination of two independent methodological frameworks, here ENMs and

Bayesian biogeographic inference, can provide a more complete evolutionary scenario, which otherwise could not be inferred by each method alone. Additionally, we introduce a novel approach to select the truncation threshold in ENM presence/absence models that is based on information provided by phylogenetic biogeographic inference.

Acknowledgments: We thank (alphabetically) Alberto Herrero, Sanna Olsson, Lisa Pokorny, Ricarda Riina for help with training and DNA sequencing; Nicola de Maio for his advice with BASTA; Gonzalo Nieto Feliner and Ricarda Riina for proof reading. VC was funded by Ministerio de Economía y Competitividad MINECO Ph.D. (FPI) Fellowship BES-2013-065389, supervised by IS. IS was funded by the Spanish Government and European Regional Development Fund, grant CGL2015-67849-P (MINECO/FEDER).

Author contributions: IS and VC designed the study. MM and VC collected and sourced samples. VC analysed the data with help from IS. VC and IS co-wrote the manuscript with significant contributions from MM.

REFERENCES

Allouche O., Tsoar A., Kadmon R. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). Journal of Applied Ecology. 43:1223-1232.

Antonelli A., Sanmartín, I. 2011. Mass extinction, gradual cooling, or rapid radiation: Reconstructing the spatiotemporal evolution of the ancient angiosperm genus *Hedyosmum* (Chloranthaceae) using empirical and simulated approaches. Systematic Biology. 60:596-615.

Araújo M.B., New M. 2007. Ensemble forecasting of species distributions. Trends in Ecology and Evolution. 22:42-47.

Araújo M.B., Ferri-Yáñez F., Bozinovic F., Marquet P.A., Valladares F., Chown S.L. 2013. Heat freezes niche evolution Ecology Letters. 16:1206-1219.

Barres L., Sanmartín I., Anderson C.L., Susanna A., Buerki S., Galbany-Casals, M., Vilatersana R. 2013. Reconstructing the evolution and biogeographic history of tribe *Cardueae* (Compositae). American Journal of Botany. 100:867- 882.

Basille M., Calenge C., Marboutin E., Andersen R., Gaillard, J.M. 2008. Assessing habitat selection using multivariate statistics: Some refinements of the ecological-niche factor analysis. Ecological Modelling. 211: 233-240.

Beerling D.J., Berner R.A., Mackenzie F.T., Harfoot M.B., Pyle J.A. 2009. Methane and the CH4 related greenhouse effect over the past 400 million years. American Journal of Science. 309:97-113.

Beerling D.J., Taylor L.L., Bradshaw C.D.C., Lunt D.J., Valdes P.J., Banwart S.A., Pagani M., and Leake J.R. 2012. Ecosystem CO_2 starvation and terrestrial silicate weathering: mechanisms and global-scale quantification during the late Miocene. Journal of Ecology. 100: 31-41.

Bouckaert R., Heled J., Kühnert D., Vaughan T., Wu C-H., Xie D., Suchard M.A., Rambaut A., Drummond A.J. 2014. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. PLoS Computational Biology. 10: e1003537.

Bradshaw C.D., Lunt D.J., Flecker R., Salzmann U., Pound M.J., Haywood A.M., Eronen J.T. 2012. The relative roles of CO_2 and palaeogeography in determining Late Miocene climate: Results from a terrestrial model-data comparison. Climate of the Past Discussions. 8:1257-1258.

Carboni M., Guéguen M., Barros C., Georges D., Boulangeat I., Douzet R., Dullinger S., Klonner G.,

van Kleunen M., Essl F., Bossdorf O., Hæuser E., Talluto M.V., Moser D., Block S., Conti L., Dullinger I., Münkemüller T., Thuiller W. 2018. Simulating plant invasion dynamics in mountain ecosystems under global change scenarios. Global Change Biology. 24: e289-e302.

Chau T.H., Rahfeldt W.A., Olmstead R.G. 2018. Comparison of taxon-specific versus general locus sets for targeted sequence capture in plant phylogenomics. Applications in Plant Sciences. 6:e1032.

Chefaoui R.M., Lobo J.M. 2008. Assessing the effects of pseudo-absences on predictive distribution model performance. Ecological Modelling. 210:478-486.

Clark J.D., Hayes S.G., Pledger J.M. 1998. A female black bear denning habitat model using a geographic information system. Ursus. 10:181-185.

Cornuault J., Sanmartín I. Comparative phylogeography: the origin of variation in dispersal patterns. International Biogeographic Society Meeting, Malaga. 8th-12th January 2019.

Couvreur T.L.P., Kissling W.D., Condamine F.L., Svenning J-C, Rowe N.P., Baker W.J. 2014. Global diversification of a tropical plant growth form: environmental correlates and historical contingencies in climbing palms. Frontiers in Genetics. 5:doi: 10.3389/fgene.2014.00452.

Craven P., Craven D. 2000. The flora of the Brandberg, Namibia. Cimbebasia Memoir. 9: 49-67.

Darriba D., Taboada G.L., Doallo R., Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. Nature Methods. 9:doi.org/10.1038/nmeth.2109.

De Maio N., Wu C-H., O'Reilly K.M., Wilson D. 2015. New routes to phylogeography: A Bayesian structured coalescent approximation. PLoS Genetics. 11:e1005421.

deMenocal P.B. 2004. African climate change and faunal evolution during the Pliocene-Pleistocene. Earth and Planetary Science Letters. 220:3-24.

Diniz-Filho J.A.F., Bini L.M. 2008. Macroecology, global change and the shadow of forgotten ancestors. Global Ecology and Biogeography. 17:11-17.

Diniz-Filho J.A.F., Ferro V.G., Santos T., Nabout J.C., Dobrovolski R., De Marco Jr. P. 2010. The three phases of the ensemble forecasting of niche models: geographic range and shifts in climatically suitable areas of *Utetheisa ornatrix* (Lepidoptera, Arctiidae). Revista Brasileira de Entomologia. 54:339-349.

Domina G., Marino P., Spadaro V., Raimondo F.M. 2012. Vascular flora evolution in the major Mediterranean islands. Biodiversity Journal. 3:337-342.

Dorn A., Musilová Z., Platzer M., Reichwald K., Cellerino A. 2014. The strange case of East African annual fishes: aridification correlates with diversification for a savannah aquatic group. BMC Evolutionary Biology. 14:210-242.

Drummond A.J., Suchard M.A., Xie D., Rambaut, A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular Biology and Evolution. 29:1969-1973.

Du Y., Mao L., Queenborough S.A., Freckleton R.P., Chen B., Ma K. 2015. Phylogenetic constraints and trait correlates of flowering phenology in the angiosperm flora of China. Global Ecology and Biogeography. 24:928-938.

Edgar R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research. 32:1792-1797.

Engler R., Guisan A., Rechsteiner L. 2004. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. Journal of Applied Ecology. 41:263-274.

Emerson B. C. 2003. Genes, geology and biodiversity: faunal and floral diversity on the island of Gran Canaria. Animal Biodiversity and Conservation. 26:9-20.

Espíndola A., Pellissier L., Maiorano L., Hordijk W., Guisan A., Alvarez N. 2012. Predicting present and future intra-specific genetic structure through niche hindcasting across 24 millennia. Ecology Letters. 15:649-657.

Felsenstein J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. Journal of Molecular Evolution. 17:368-376.

Fick S.E., Hijmans R.J. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global

land areas. International Journal of Climatology. 37:4302-4315.

Fielding A.H., Bell J.F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation. 24:38-49.

Galley C., Bytebier B., Bellstedt D.U., Linder H.P. 2007. The Cape element in the Afrotemperate flora: from Cape to Cairo? Proceedings of the Royal Society B: Biological Sciences. 274:535-543.

García-Aloy S., Sanmartín I., Kadereit G., Vitales D., Millanes A.M., Roquet C., Vargas P., Alarcón M., Aldasoro J.J. 2017. Opposite trends in the genus Monsonia (Geraniaceae): specialization in the African deserts and range expansions throughout eastern Africa. Scientific Reports. 7:10.1038/s41598-017-09834-6.

Giorgi F., Lionello P. 2008. Climate change projections for the Mediterranean region. Global Planetary Change. 63:90-104.

Griffin D.L. 2002. Aridity and humidity: two aspects of the late Miocene climate of North Africa and the Mediterranean. Palaeogeography, Palaeoclimatology, Palaeoecology. 182:65-91.

Hæuser E., Dawson W., Thuiller W., Dullinger S., Block S., Bossdorf O., Carboni M., Conti L., Dullinger I., Essi F., Klonner G., Moser D., Münkemüller T., Parepa M., Talluto M.V., Kreft H., Pergi J., Pyšek P., Weigelt P., Winter M., Hermy M., Van der Veken S., Roquet C., van Kleunen M. 2018. European ornamental garden flora as an invasive debt under climate change. Journal of Applied Ecology. 55:2386-2395.

Hengl T., Sierdsema H., Radović A., Dilo A. 2009. Spatial prediction of species' distributions from occurrence-only records: combining point pattern analysis, ENFA and regression-kriging. Ecological Modelling. 220:3499-2511.

Hewitt G.M. 2004. Genetic consequences of climatic oscillations in the Quaternary. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences. 359:183-195.

Hilliard O. 1994. The Manuleae: a Tribe of Scrophulariaceae. Edinburgh University Press, Edinburgh.

Höhna S., Landis M.J., Heath T.A., Boussau B., Lartillot N., Moore B.R., Huelsenbeck J.P., Ronquest F. 2016. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification Langauge. Systematic Biology. 65:726-736.

Jiménez-Valverde A., Lobo J. M. 2007. Threshold criteria for conversion of probability of species presence to either-or presence-absence. Acta Oecologica. 31:361-369.

Kornhall P., Heidari N., Bremer B. 2001. Selagineae and Manuleeae, two tribes or one? Phylogenetic studies in the Scrophulariaceae. Plant Systematics and Evolution. 228:199-218.

Landis M.J., Matzke N.J., Moore B.R., Huelsenbeck J.P. 2013. Bayesian analysis of biogeography when the number of areas is large. Systematic Biology. 62:789–804.

Landis J.B., Soltis D.E., Li Z., Marx H.E., Barker M.S., Tank D.C., Soltis P.S. 2018. Impact of whole-genome duplication events on diversification rates in angiosperms. American Journal of Botany. 105:348-363.

Lemey P., Rambaut A., Drummond A.J., Suchard M.A. 2009. Bayesian phylogeography finds its roots. PLoS Computational Biology. 5:e1000520. doi:10.1371

Liu C., Berry P.M., Dawson T.P., Pearson R.G. 2005. Selecting thresholds of occurrence in the predictions of species distributions. Ecography. 28:385-393.

Liu H., Li S., Ugolini A., Momtazi F., Hou Z. 2018. Tethyan closure drove tropical marine biodiversity: Vicariant diversification of intertidal crustaceans. Journal of Biogeography. 45:941-951.

Lobo J.M., Jiménez-Valverde A., Real R. 2008. AUC: a misleading measure of the performance of predictive distribution models. Global Ecology and Biogeography. 17:145-151.

Magallón S., Gómez-Acevedo S., Sánchez-Reyes L.L., Hernández-Hernández T. 2015. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. New Phytologist. 207:437-453.

Mairal M., Pokorny L., Aldasoro J.J., Alarcón M., Sanmartín I. 2015a. Ancient vicariance and climate-driven extinction explain continental-wide

disjunctions in Africa: the case of the Rand Flora genus *Canarina* (Campanulaceae). Molecular Evolution. 24:1335-1354.

Mairal M., Sanmartín I., Aldasoro J.J., Culshaw V., Manolopoulou I., Alarcón M. 2015b. Palaeo-islands as refugia and sources of genetic diversity within volcanic archipelagos: the case of the widespread endemic *Canarina canariensis* (Campanulaceae). Molecular Ecology. 24:3944-3963.

Mairal M., Sanmartín I., Pellissier L. 2017. Lineage-specific climatic niche drives the tempo of vicariance in the Rand Flora. Journal of Biogeography. 44:doi: 10.1111/jbi.12930.

Mairal M., Caujapé-Castells J., Pellissier L., Jaén-Molina R., Álvarez N., Heuertz M., Sanmartín I. 2018. A tale of two forests: ongoing aridification drives population decline and genetic diversity loss at continental scale in Afro-Macaronesian evergreen-forest archipelago endemics. Annals of botany. 122:1005-1017.

Martínez-Meyer, E., Peterson, A. T., Hargrove, W. W. 2004. "Ecological niches as stable distributional constraints on mammal species, with implications for Pleistocene extinctions and climate change projections for biodiversity" Global Ecology and Biogeography 13: pp 305-314.
Martínez-Meyer E., Peterson A.T. 2006. Conservatism of ecological niche characteristics in North American plant species over the Pleistocene-to-recent transition. Journal of Biogeography. 33:1779-1789.

Meseguer A.S., Lobo J.M., Ree R., Beerline D.J., Sanmartín I. 2015. Integrating fossils, phylogenies, and niche models into biogeography to reveal ancient evolutionary history: The case of *Hypericum* (Hypericaceae). Systematic Biology. 64:215–232.

Meseguer A.S., Lobo J.M., Cornuault J., Beerling D., Ruhfel B.R., Davis C.C., Jousselin E., Sanmartín I. 2018. Reconstructing deep-time palaeoclimate legacies in the clusioid Malpighiales unveils their role in the evolution and extinction of the boreotropical flora. Global Ecology and Biogeography. 27:616-628.

Metcalf J.L., Prost S., Nogués-Bravo D., DeChaine E.G., Anderson C., Batra P., Araújo M.B., Cooper A., Guralnick R.P. 2014. Integrating multiple lines of evidence into historical biogeography hypothesis testing: A *Bison bison* case study. Proceedings of the Royal Society B. 281:20132782.

Monserud R.A., Leemans R. 1992. Comparing global vegetation maps with the Kappa statistic. Ecological Modelling. 62:275-293.

Müller K. 2005. SeqState - primer design and sequence statistics for phylogenetic DNA data sets. Applied Bioinformatics. 4:65-69.

Nenzén H.K., Araújo M.B. 2011. Choice of threshold alters projections of species range shifts under climate change. Ecological Modelling. 222:3346-3354.

Novak T., Tkavc T., Kuntner M., Arnett A.E., Delakorda S.L., Perc M., Janžekovič F. 2010. Niche partitioning in orbweaving spiders *Meta menardi* and *Metellina merianae* (Tetragnathidea). Acta Oecologica. 36:522-529.

Oxelman B., Kornhall P., Olmstead R.G., Bremer B. 2005. Further disintegration of Scrophulariaceae. TAXON. 54:411-425.

Pearson R.G., Raxworthy C.J., Nakamura M., Peterson A.T. 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. Journal of Biogeography. 34:102-117.

Pellissier L., Bråthen K.A., Vittoz P., Yoccoz N.G., Dubuis A., Meier E.S., Zimmermann N.E., Randin C.F., Thuiller W., Garraud L., Van Es J., Guisan A. 2013. Thermal niches are more conserved at cold than warm limits in arctic-alpine plant species. Global Ecology and Biogeography 22:933-941.

Peterson A.T. 2006. Use and requirements of ecological niche models and related distributional models. Biodiversity Informatics. 3:59-72.

Pirie M., Kandziora M.D., Nürk N.M., Le Maitre N.C., Mugrabide Kuppler A.L., Gehrke B., Oliver E.G.H., Bellstedt D.U. 2018. Leaps and bounds: geographical and ecological distance constrained the colonisation of the Afrotemperate by *Erica*. BioRxiv. 290791:doi: 10.1101/290791.

Pokorny L., Oliván G., Shaw A.J. 2011. Phylogeographic patterns in two southern hemisphere species of *Calyptrochaeta*

(Daltoniaceae, Bryophyta). Systematic Botany. 36:542-553.

Pokorny L., Riina R., Mairal M., Meseguer A.S., Culshaw V., Cendoya J., Serrano M., Carbajal, R., Ortiz S., Heuertz M., Sanmartín I. 2015. Living of the edge: timing of Rand Flora disjunctions congruent with ongoing aridification in Africa. Frontiers in Genetics. 6:doi: 10.3389/fgene.2015.00154.

Popp M., Gizaw A., Nemomissa S., Suda J., Brochmann C. 2008. Colonization and diversification in the African 'sky islands' by *Eurasian Lychnis L*. (Caryophyllaceae). Journal of Biogeography. 35:1016-1029.

Popp M., Mirré V., Brochmann C. 2011. A single Mid-Pleistocene long-distance dispersal by a bird can explain the extreme bipolar disjunction in crowberries (Empetrum). Proceedings of the National Academy of Sciences. 108:6520-6525.

Rambaut A., Drummond A.J., Xie D., Baele G., Suchard M.A. 2018. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. Systematic Biology. syy032. doi:10.1093/sysbio/syy032.

Ree R.H., Moore B.R., Webb C.O., Donoghue M.J. 2005. A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. Evolution. 59:2299-2311.

Ree R.H., Sanmartín I. 2018. Conceptual and statistical problems with the DEC+J model of founder-event speciation and its comparison with DEC via model selection. Journal of Biogeography. 45:741-749.

Ree R.H., Smith S.A. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. Systematic Biology. 57:4-14.

Richards C.L., Carstens B.C., Knowles L.L. 2007. Distribution modelling and statistical phylogeography: an integrative framework for generating and testing alternative biogeographical hypotheses. Journal of Biogeography. 34:1833-1845.

Romdal T.S., Araújo M.B., Rahbek C. 2013. Life on a tropical planet: niche conservatism and the global diversity gradient. Global Ecology Biogeography. 22:344-350.

Ronquist F., Teslenko M., van Der Mark P., Ayres D.L., Darling L., Höhna S., Larget B., Liu L., Suchard M.A., Hueslsenbeck J.P. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology. 61:539-542.

Ronquist F., Sanmartín, I. 2011. Phylogenetic methods in biogeography. Annual Review of Ecology, Evolution, and Systematics. 42:441-464.

Roquet C., Sanmartín I., Garcia-Jacas N., Sáez L., Susanna A., Wikström N., Aldasoro, J.J. 2009. Reconstructing the history of Campanulaceae with a Bayesian approach to molecular dating and dispersal–vicariance analyses. Molecular Phylogenetics and Evolution. 52:575-587.

Sanmartín I., van Der Mark P., Ronquist F. 2008. Inferring dispersal: A Bayesian approach to phylogeny-based island biogeography, with special reference to the Canary Islands Journal of Biogeography. 35:428-449.

Sanmartín I., Anderson C.L., Alarcon M., Ronquist F., Aldasoro J.J. 2010. Bayesian island biogeography in a continental setting: The Rand Flora case. Biology Letters. 6:703-707.

Sanmartín I., Meseguer A.S. 2016. Extinction in phylogenetics and biogeography: from timetrees to patterns of biotic assemblage. Frontiers in Genetics. 7:35.

Saupe E.E., Barve N., Owens H.L., Cooper J.C., Hosner P.A., Peterson A.T. 2017. Reconstructing ecological niche evolution when niches are incompletely characterized. Systematic Biology. 67:428-438.

Senut B., Pickford M., Ségalen L. 2009. Neogene desertification of Africa. Comptes Rendus Geoscience. 341:591-602.

Sepulchre P., Ramstein G., Fluteau F., Schuster M., Tiercelin J.J., Brunet M. 2006. Tectonic uplift and Eastern Africa aridification. Science. 313:1419-1423.

Simmons M.P., Ochoterena H. 2000. Gaps as characters in sequence-based phylogenetic analysis. Systematic Biology. 49:369-381.

Smith S.A., Donoghue M.J. 2010. Combining historical biogeography with niche modeling in the *Caprifolium* clade of *Lonicera* (Caprifoliaceae, Dipsacales). Systematic Biology. 59:322-341.

Stadler T. 2009. On incomplete sampling under birth-death models and connections to the sampling-based coalescent. Journal of Theoretical Biology. 261:58-66.

Svenning J-C., Eiserhardt W.L., Normand S., Ordonez A., Sandel B. 2015. The influence of palaeoclimate on present-day patterns in biodiversity and ecosystems. Annual Review of Ecology, Evolution, and Systematics. 46:551-572.

Tagliacollo V.A., Roxo F.F., Duke-Sylvester S.M., Oliveira C., Albert J.S. 2015. Biogeographical signature of river capture events in Amazonian lowlands. Journal of Biogeography. 42:2349-2362.

Thiv M., Thulin M., Hjertson M., Kropf M., Linder H.P. 2010. Evidence for a vicariant origin of Macaronesian–Eritreo/Arabian disjunctions in *Campylanthus* Roth (Plantaginaceae). Molecular Phylogenetics and Evolution. 54:607-616.

Thuiller W., Lavorel S., Sykes M.T., Araújo M.B. 2006. Using niche-based modelling to assess the impact of climate change on tree functional diversity in Europe. Diversity and Distributions. 12:49-60.

Trauth M.H., Larrasoaña J.C., Mudelsee M. 2009. Trends, rhythms and events in Plio-Pleistocene African climate. Quaternary Science Reviews. 28:399–411.

Vaughan T., Drummond P., Drummond, A. 2012. Within-host demographic fluctuations and correlations in early retroviral infection. Journal of Theoretical Biology. 295:86-99.

Villaverde T., Pokorny L., Olsson S., Rincón-Barrado M., Johnson M.G., Gardner E.M., Wickett N.J., Molero J., Riina R., Sanmartín I. 2018. Bridging the micro- and macroevolutionary levels in phylogenomics: Hyb-Seq solves relationships from populations to species and above. New Phytologist. 220:636-650.

Villaverde T., Rincón M., Sanmartín I. Rare and widespread: target-enrichment approaches reconstruct the origins of the Rand Flora pattern in *Camptoloma* (Scrophulariaceae). International Biogeographic Society Meeting, Malaga. 8th-12th January 2019.

Waldron A. 2010. Lineages that cheat death: surviving the squeeze on range size. Evolution. 64: 2278-2292.

Wiens J.J. 2004. Speciation and ecology revisited: phylogenetic niche conservatism and the origin of species. Evolution. 58:193–197.

Willis K.J., MacDonald G.M. 2011. Long-term ecological records and their relevance to climate change predictions for a warmer world. Annual Review of Ecology, Evolution, and Systematics. 42:267–287.

Zachos J., Pagani M., Sloan L., Thomas E., Bullups K. 2001. Trends, rhythms, and aberrations in global climate 65 Ma to Present. SCIENCE. 292:686-693.

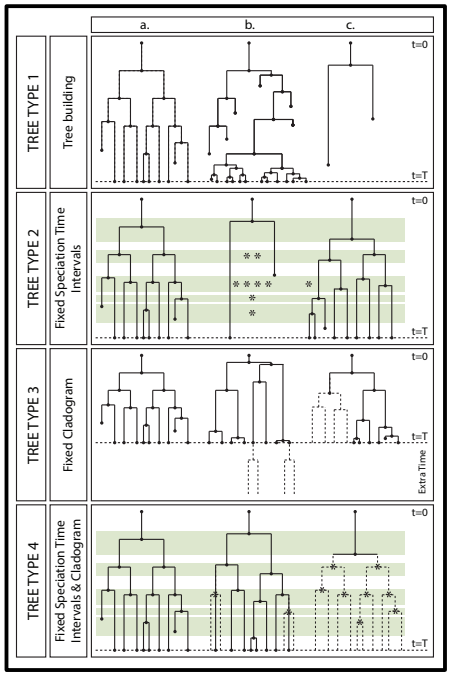
Zachos J.C., Gerald R.D., Zeebe R.E. 2008. An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. NATURE. 451:279-283.

CHAPTER 3

Evolutionary macroecology: Incorporating phylogenetic information more explicitly into process-based, forward time, simulation models

This chapter has been realised in collaboration with Thiago F. L. V. B. Rangel (LETS, Universidade Federal de Goiás, Brasil) and Isabel Sanmartín (Department of Biodiversity and Conservation, Real Jardín Botánico, CSIC, Spain).

This chapter is planned to be submitted to the peer-review journal *Ecography*.



CHAPTER 3

A Spatially Explicit Computer Simulation Model to Predict the Origin, Spreading and Extinction of Geographic Ranges Using Environmental and Evolutionary History

Victoria Culshaw, Thiago F.L.V.B. Rangel, Isabel Sanmartín

ABSTRACT

A conceptual and mechanistic approach for bridging the fields of macroecology (ecological biogeography) and historical biogeography has been a long-term aim in Evolutionary Biology. Such a bridge could increase our understanding on the processes governing the spatial and temporal generation of biodiversity distribution patterns. This aim has been recently approached from the perspective of evolutionary biogeographic inferential statistical models, within a maximum likelihood or Bayesian framework, which incorporates the contribution of environmental factors as scaling parameters. Here, we describe a new approach that builds on a spatially-explicit, forward-time, computer simulation ("automat") model. The model sets a series of rules by which speciation, extinction, and dispersal of lineages are governed within an environmentally heterogeneous, two-dimensional gridded landscape. Unlike some previous approaches, niche conservatism is assumed but the model allows for environmental conditions to vary both spatially and temporally, by letting the model run over three time-series of actual palaeoclimate data, spanning the last 20 million years of geological history. Speciation is governed by a global speciation rate, whereas the background extinction rate is made dependent on abiotic (palaeoenvironmental conditions) and biotic (species density per grid cell) factors, hence giving a local background extinction rate. Also, we propose a novel mechanistic approach in which species are not the result of unique, independent events but linked through evolutionary history from a single evolutionary origin. We set different rules to generate the resulting phylogenies to test different factors (time, environment) that govern the inheritance of range distributions. Dispersal follows a simple Poisson kernel model, with higher probability of migration to contiguous grid cells and rare long-distance movements to distant cells. We also describe ways in which the presence of temporal dispersal barriers could modify the resultant spatial patterns. Evaluation of model accuracy and fit is based on comparison of simulated spatial patterns with observed, empirical ones. We use statistic dependent variables such as the spatial distribution of species in the landscape, species' geographic range size and location, and the shape of the resultant phylogenies. Finally, we propose that this spatially-explicit simulation model could be used to evaluate the role played by niche conservatism, ecological vicariance and climatic-driven extinction in the generation of disjunct continental patterns, such as the presence of sister-lineages in numerous families of angiosperms forming a ring in the continental margins of the African continent, the Rand Flora pattern.

Running-title: Macroecological Simulation Model with Evolutionary History

Keywords: Ancient Climate Change, Computational Simulations, Ecological Niche Model, Evolutionary History, Extinction, Macroecology, Mechanistic Approach, Rand Flora

INTRODUCTION

Understanding biodiversity patterns – why some biotas are more or less diverse than others and how local or regional biotas became assembled– has become a pressing goal in biodiversity research in light of the Anthropocene biodiversity crisis (up to 25% of

extant species will become extinct in the next decades; IPBES, https://ipbes.net) and ongoing global climate change scenario (linked to human activities). Global biodiversity patterns, such as the Latitudinal Diversity Gradient (LDG) –the observed general increase in species richness or biodiversity from the poles to the tropics (Fischer 1960; Pianka 1966; Stevens 1989)– have traditionally been explained in relation to shifts in environmental variables such as temperature or precipitation (Hillebrand 2004; Kreft and Jetz 2007). There is, though, a current understanding that environmental variables cannot by themselves increase or decrease local or regional species richness; evolutionary processes such as dispersal, extinction, or speciation must be taken into account (Qian and Ricklefs 2004; Wiens and Donoghue 2004; Jablonski et al. 2008; Mittelback et al. 2007). Thus, understanding the mechanistic, causal factors driving speciation, dispersal, and extinction rates in relation to fluctuating spatially and temporal environmental (climatic) conditions should allow us to comprehend why some biotas are richer or more species-poor than others or how global biodiversity patterns have been shaped over time. For example, it has been hypothesised that if niche conservatism prevails over niche evolution in regions of spatially heterogeneous and temporally fluctuating climate, diversification will occur predominantly by a process of range fragmentation, caused by the inability of species to adapt to changing environmental conditions in portions of the ancestral range, in what has been termed the process of "ecological vicariance" (Wiens 2004; Wiens and Donoghue 2004; Rangel et al. 2007).

These questions have been typically approached from two different perspectives. One of them is macroecological, by linking measures of diversity and spatial patterns of species richness to environmental (often, climatic) variables through "curve-fitting" correlative approaches, such as generalised linear modelling techniques (GLM; Hawkins et al. 2003; Rahbek et al. 2007; Chefaoui and Lobo 2007; Mairal et al. 2017), or through more complex spatial Ecological Niche Models (ENM) and Species Distribution Models (SDM, see Owens et al. 2011 for distinction between ENM and species distribution models, SDMs, from here we consider ENM only; Thuiller 2003; Araujo et al. 2005; Thuiller et al. 2006; Lobo et al. 2008; Hengl et al. 2009; Culshaw et al. in rev.). ENMs are a group of distances or models that are used to describe and predict climatic habitat suitability species' geographical ranges from occurrence, and if possible absence data, e.g. ecological niche factor analysis (ENFA; Hirzel et al. 2002; Basille et al. 2008; Hengl et al. 2009) or Mahalanobis distances (MD, Clark et al. 1998). ENMs for individual species can be overlaid to extract general patterns on climatic habitat suitability (Culshaw et al. in rev.) These models have later been extended with the addition of

other parameters, such as species interactions, non-climatic predictors, or spatial autocorrelation terms (Araújo and Luoto 2007; Dormann et al. 2007; De Marco et al. 2008; Guisan and Rahbek 2011). Currie et al. (1999) criticised curve-fitting approaches because they often do not provide quantitative predictions and because they "implicitly" assume that species richness patterns depend only on environmental variables, ignoring the contribution of historical and evolutionary events to such patterns. These critiques can also be extended to ENMs: overlapping of potential or realised species ranges in a gridded map does not provide a causal explanation behind the observed pattern.

The second approach to tackle the origin and maintenance of spatial patterns of diversity is using methods stemming from the fields of phylogenetics and historical biogeography (Sanmartín 2014). These methods use phylogenetic data (evolutionary relationships) and divergence time estimates, together with species occurrence data, to infer the distribution range of a species' ancestors and the rate of evolutionary processes like speciation, extinction, and dispersal (Wiley 1981; Brooks and McLennan 1991; Ronquist and Sanmartín 2011; Sanmartín 2014). Current biogeographical inferential approaches are based on probabilistic models of range evolution (continuous-time discrete-state Markov chain models) and use the statistical maximum likelihood framework or Bayesian Inference to estimate ancestral states and parameter values. Unlike in ENMs, areas in biogeography are often defined in abstract terms, without an explicit spatial or environmental component. However, as some biogeographers have noted (Brooks 1990; Cracraft 1994; Sanmartín et al. 2001; Wiens and Donoghue 2004; Meseguer et al. 2015), the environmental (climatic) preferences of a species plays a strong role in constraining evolutionary and biogeographical processes such as the rate of ecological speciation, local background extinction and geographic expansion, and therefore should be considered in the inference of spatial patterns of species richness.

Understanding the mechanistic basis of broad-scale biodiversity patterns formed by persistence, adaptation, extinction or geographical range shifts remains the holy grail of modern biogeography and macroecology (Judson 1994; Willig et al. 2003; Gotelli et al. 2009). In the early days of these disciplines (XIX century and early XX century), phylogeny and biogeography were closely interlinked in the works of taxonomists and naturalists (Cox, Moore and Ladle. Biogeography: An Ecological and Evolutionary Approach, 10th edition, 2016). However, for over the last 40 years, the field of biogeography has been mainly concerned with building phylogenies and inferring the spatial-temporal evolution of lineages, with ecology deemed moot (Wiley 1981; Brooks and

McLennan 1991; Ree and Smith 2008; Sanmartín et al. 2008; Lemey et al. 2009); conversely, macroecology has been slow in integrating the evolutionary perspective (Hawkins et al. 2003). Yet, historical biogeography and ecology have much to learn from one another (Wiens and Donoghue (2004); Hawkins et al. (2005); Culshaw et al. (in rev.), and an approach that bridges these two fields into a single theoretical and analytical framework will be crucial in exploring the relationship between ecological and evolutionary processes in the shaping of Earth's species geographic ranges (Ricklefs 2006).

Recently, there have been some attempts to create this bridge. Studies such as Evans et al. (2009), Richards et al. (2007) or Meseguer et al. (2015) have championed the integration of phylogenetic and biogeographical inference with ENM frameworks. Richards et al. (2007) used hindcasted ENMs to inform coalescent simulations and test alternative coalescent hypotheses on population spatial dynamics. Evans et al. (2009) employed ENMs to quantify the climatic disparity among taxa and obtain an understanding of the degree along which niches evolved, and used this information as a secondary calibration to predict niche occupancy through the timeline of a dated phylogenetic tree. Meseguer et al. (2015) combined fossil record information with presence only data to infer the climatic tolerances of ancestral lineages and generate hindcasted ENMs. Subsequently, they used these models, as well as the spatial information provided by fossils, to inform probabilistic biogeographic inference about the existence of temporal climatically suitable corridors and climatic dispersal barriers. We recently proposed an approach (Culshaw et al. in rev) to combine the biogeographic and ENM frameworks as complementary, independent sources of evidence, allowing the user to integrate the geographical information without sacrificing the evolutionary information, and vice versa. Hindcasted ENMs were used to inform biogeographic inference along long branches lacking cladogenetic events, while inferred ancestral ranges were used to select the truncation threshold needed to convert continuous habitat suitability values returned by the ENM into presence-absence predictions.

Though these integrative/combined approaches can improve the accuracy of the inferred biogeographic scenarios or the parameter estimates (Meseguer et al. 2015; Culshaw et al. in rev), they are more geared towards understanding the spatio-temporal history of particular lineages and less to the testing of hypotheses on macroecological patterns. Biogeographic and ENM approaches do not allow us to examine the causes behind an observed correlation between species distribution patterns and environmental factors, or what this relationship would be like under small variations in the primary data. Knowing

how stable a statistical relationship is, provides a better understanding of its predictive power in forecasting future scenarios. A mechanistic understanding of species distribution patterns requires modelling the processes by which these patterns are formed (Gotelli et al. 2009).

A third avenue to combine biogeography and macroecology is based on the use of spatially explicit, forward-time computational general simulation models (GSM, Rangel et al. 2006, 2007; Rabehk et al. 2007; Colwell et al. 2009; Gotelli et al. 2009), sometimes coined as "automat" or "in silico" models. In its simplest form, GSMs are mechanistic models that are governed by a set of rules for location; probability and mechanism of speciation; the inheritance of niche characteristics by each new species from its immediate ancestor; and the species ability to disperse to and successfully colonise new grid cells based on their environmental characteristics (Rangel et al. 2007; Gotelli et al. 2009). These sets of rules describe how a species may speciate, disperse and become extinct in an environmentally heterogeneous landscape (represented as a gridded domain; Gotelli et al. 2009). Typically, these models are probabilistic or stochastic (an example exception would be Hassell et al. 1991), and are run for a given number of time steps t or until a specific condition is met (e.g. until a particular number of species ranges are simulated or until a balance between speciation and extinction is achieved) and can accommodate for contemporary, past or future climates, evolutionary and historical forces, and geometric constraints. These models build frameworks to investigate hypotheses about the relative influence on species richness played by geometric dispersal constraints, environmental factors and historical processes (Rabhek et al. 2007; Gotelli et al. 2009), making these models more than suitable for exploring the stability of an empirical relationship. See Gotelli et al. (2009) for an excellent review on macroecology simulation models.

Spatially explicit general simulation models (GSM)

Exploring the causes of spatial variation in species distributions through computational general simulation models (GSM) has been the subject of a rich literature. The GSM relies only on two data layers: a gridded map of a region or biogeographic domain with cells of a certain resolution; and the environmental variables measured on each grid cell. Unlike ENMs, GSMs do not require species occurrence data. Most GSMs include also three components or control knobs for simulating species richness patterns on the gridded domain: dispersal limitation, environmental gradients, and evolutionary events (Gotelli et al. 2009).

Dispersal Limitation: Levins (1969) was one of the first models that explored the introduction of spatial parameters into population models. Levins's model describes a

gridded landscape in which present in each cell is an unstable population system. Here a population becomes extinct within a specified number of time steps. However, when several populations are linked together through a spatial parameter modelling dispersal, the populations only face temporarily extinction as they can be re-colonised by another population, hence forming a stable "on our own we are weak, together we are strong" system, provided that the population models' extinctions are in a state of asynchrony to one another. Levins's model was very simple, assuming the landscape to be homogenous, and dispersal to be limitless to any cell within the gridded domain. Also the model only recorded whether a cell was occupied or not, i.e. number of individuals of the population within the cell was disregarded. This model of linked populations became known as the Levins metapopulation model, and it is widely regarded as the null model in exploring spatial patterns through GSMs (Gotelli et al. 2009). Further advances from these first models have included the "spreading dye model", which imposed limitations on how a species may disperse through the landscape from its original cell: i.e. dispersal is determined by distance to the adjacent grid cells (e.g. Hassell et al. 1991; Jetz and Rahbek 2001; Connolly 2005).

Environmental gradients are typically based on temperature and precipitation measurements for each cell in the domain, and usually represent present-day climate data, and, increasingly popularly, palaeoclimate reconstructions (Meseguer et al. 2015; Culshaw et al. in rev). With the use of ENMs, these environmental gradients can be converted into species-specific, habitat suitability values, where one of these suitability values is allocated per cell in the domain. Converting environmental variables into habitat suitability variables can be advantageous. One advantage would be that habitat suitability values from the ENMs can locate the potential spatial barriers to the spatial distribution of a population through time, and the GSM can be used to see whether it is possible for the population to overcome these spatial barriers given enough time, hence effectively mechanising the ENM (Thuiller et al. 2008). A second advantage is that speciation, background extinction and colonisation success rates can be converted by the ENM habitat suitability values from being global rates that are equal across species to local and species-specific.

Including *evolutionary events* in GSMs has mainly been tackled in one of two ways: i. by specifying a priori the number of independent evolutionary events that should take place during the simulation (e.g. Grytnes 2003; Storch et al. 2006; Rahbek et al. 2007); and ii. by including some scale or logic measurement of what a species is, running the simulation for a given time, and then observing how many evolutionary events took place and how many species are present at the end of the simulation (e.g. Rangel et al 2007; Boone

2010). In the first case, each species has an independent evolutionary origin and, if the niche is considered, the niche of each species is independent from the niche of other species; in the second, all species are linked by a single evolutionary origin, an ancestral species, and the sequence of speciation events depends on the set of rules that determines the balance between niche conservation and niche evolution (Rangel et al. 2007).

Evaluation of the power of GSMs is based on comparison of simulated and observed patterns, and the ability to predict or recover system-level properties such as the LDG pattern or the observed mid-domain effect (Gotelli et al. 2009). Comparison between simulations and the observed input data can also be used to find the optimal parameter values: these are the parameter combinations that maximise the similarity between the observed and the simulated patterns, for example, in regards to species richness and/or the geographic range distribution frequency among species (Rangel et al. 2007). In fact, one advantage of GSM approaches, which resembles that of Approximate Bayesian Computation methods in evolutionary biology (Leuenberger and Wegman 2010), is their flexibility to incorporate new parameters or variables to increase the realism of the model, since it becomes unnecessary to define the mathematical functions governing parameter behaviour/relationships analytically (i.e., likelihood-free methods).

Some Limitations of GSMs

One of the main limitations of GSMs is the assumption of species equivalency. As in MacArthur and Wilson's (1967) Island Biogeography Theory, or Hubbell's (2001) Neutral Theory, rates of speciation, background extinction, and dispersal do not depend on species-specific characteristics, but are assumed similar or equal across species. Model realism could also be increased by incorporating actual palaeoclimate data and a real time scale in the simulations, rather than using an artificially fluctuating environment (Rangel et al. 2007). To consider speciation events as independent evolutionary events follows the classic ecological approach of using numbers of species or species richness to detect patterns (Jetz and Rabnek 2001; Wiens and Donoghue 2004), but does not help understand how these numbers are generated (Wiens and Donoghue 2004). Modelling evolutionary events using a scale and logic measurement of what a species is through a set of rules (Rangel et al. 2007) allows comparing the observed species richness patterns and geographic range size variation with the one obtained in the simulations, and hence to extract conclusions on the underlying processes. Yet, deciding upon this scale and logic measurement can be difficult, as the initial empirical ecological niche, the empirical heritability of the ecological niche from ancestor to descendant taxon, and an understanding of the balance

between empirical niche conservation and empirical niche evolution through time must be known. This difficulty increases the appeal of removing oneself from the real world (such as done in studies like Rangel et al. 2007), and hence remove the need for empirical information and allowing the luxury of choice as to which type of rules are set. However, with this entrance into the virtual world comes a new challenge, the lack of an evolutionary component provided by a real phylogeny, for example, the inheritance of the ancestral niche by the descendants. Setting the simulation so as to explore the relationship between a particular phylogeny, or empirical species relationship, with a distribution pattern, climate history and niche dynamics allows testing whether a similar relationship could be recreated under different environmental conditions and niche dynamics. In other words, are the relationships we find within a phylogenetic tree linked to the distribution patterning or could the same relationships have been achieved through different distribution patterns, and vice versa?

Here, we describe a novel general simulation model which proposes a different way of integrating evolutionary history in GSMs in the landscape. The model includes a dispersal kernel to navigate around temporary barriers, and information on the tree and speciation intervals among species, so that the evolutionary (speciation) events in the simulation are dependent on one another, i.e., they have a single origin. We gain the information for this new evolutionary component by including phylogenetic information into the simulation. Unlike in Rangel et al. (2007), niche conservation is assumed within our model (i.e., no information on niche evolution); hence, evolutionary events are initiated by a speciation rate rather than by the balance between niche conservatism and evolution in a fluctuating environment. Moreover, our GSM uses time series of paleoclimatic data: in particular, three paleoclimate layers representing major cooling and warming events in the last 20 million years. As in other GSMs, the model uses a pattern-oriented modelling approach to compare the fit between observed and simulated data.

A Novel General Simulation Model Introducing Evolutionary History

Model Overview

Our new model follows Wiens and Donoghue (2004) biogeographic framework, in which the geographical range of a clade is determined by i. the ancestral ecological niche of the clade; ii. the geographical dispersal origin; iii. dispersal limitation described by abiotic and biotic conditions (e.g. habitat suitability and competition); iv. the potential for niche evolution presented by the geographical range, if niche conservation is not considered), and the length of time elapsed since the origin of the clade for niche evolution (if niche conservation is not considered) and

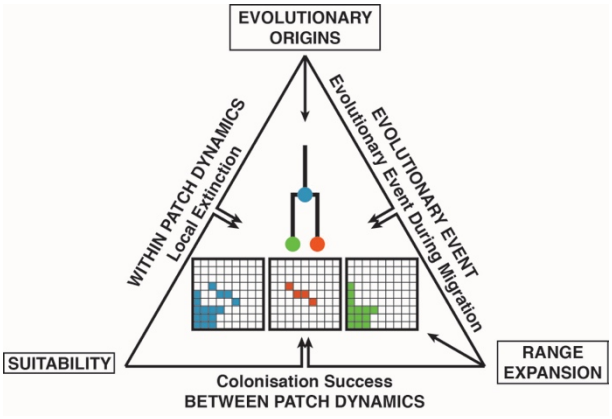


Figure 1. Model Overview. The points of the triangle represent the data layers supplied to the model. The triangle sides represent the connecting relationships between these layers. The black arrows represent the direct influence over the centre model results. Notice that "Suitability" only affects the model indirectly through "Within Patch Dynamics" and "Between Patch Dynamics".

dispersal to occur. Our spatially explicit, forward time GSM model directly addresses four of these points (i-iii and v). Point iv (the potential for niche evolution) is not addressed; since we assume niche conservatism or at least that we lack enough information to infer niche evolution. The model is built upon an n-dimensional niche space and a two-dimensional geographical, gridded cell map space. Each cell contains information on the number of species present within it, but not the number of individuals for each species, as well as its geographical coordinates, and by the local values of the same n environmental parameters that define the niche space (a representation pioneered by Pulliam 2000). The model has been written in the programming language PASCAL, and is available for distribution by demand to main author (vickycul@hotmail.com).

Figure 1 shows a scheme representing the model, showing the three components or control knobs in the model (Figure 1 the vertices of the triangle), and the relationships among these components in the model (Figure 1 the sides of the triangle). The three control knobs are the: a. Evolutionary Origins: including the rates of extinction (μ) and speciation (λ), and the time intervals between speciation events and/or the species tree topology, which provides information as to when a speciation event takes place and to whom this speciation event is related. b. Suitability: provides information on the abiotic and biotic suitability of each cell. Abiotic suitability defines how suitable the cell is for the species, given its climatic niche breadth and the measurements of abiotic variables, such as mean temperature and precipitation, in the cell. Biotic suitability refers to competition within each cell and is represented simply by the number of species present within a cell (i.e., density-dependence effects). c. Range Expansion: describes the ability of a species to disperse across the landscape. We used a model that combines the "spreading dye model", allowing local dispersal to adjacent cells within a maximum of distance M, with rare events of long-distance ("freak") dispersal to cells located beyond M.

Relationships between components in the model (Figure 1, represented by the triangle

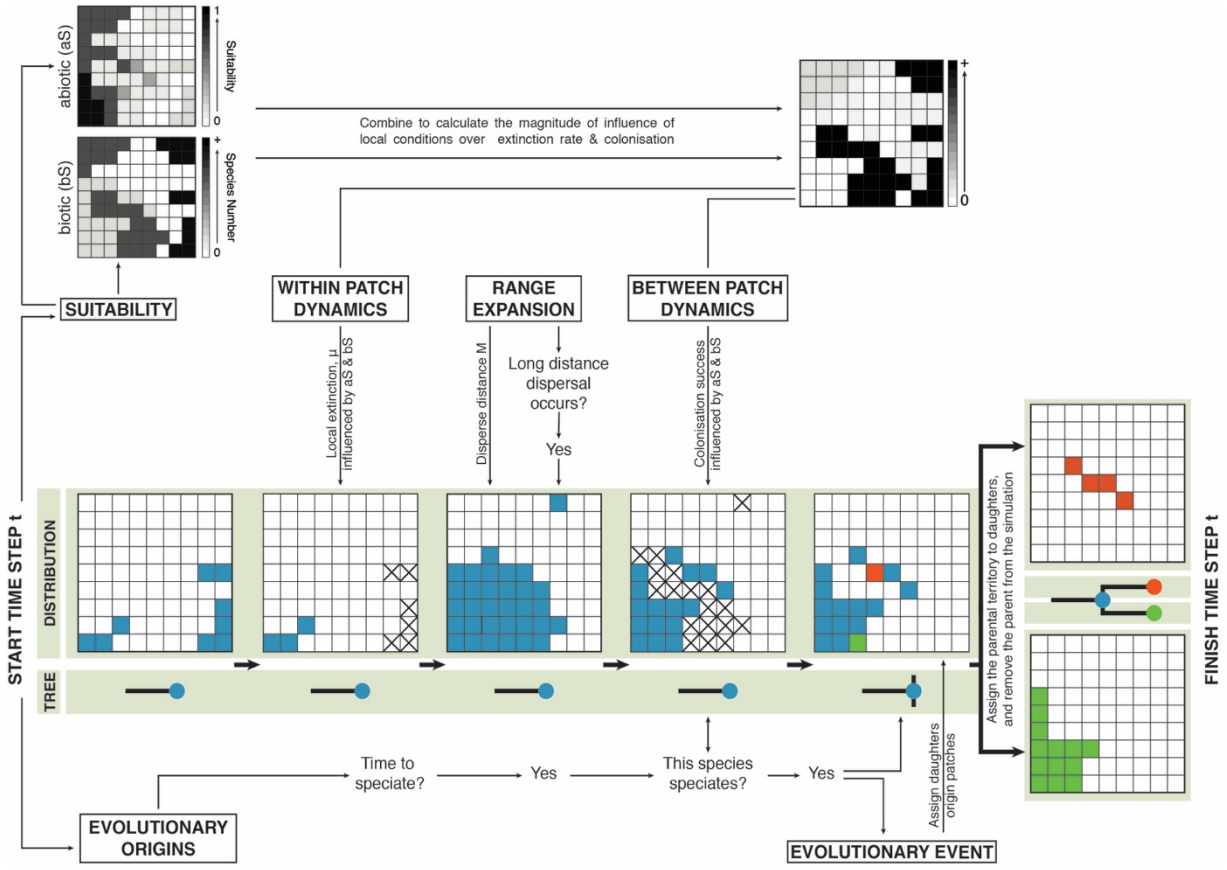


Figure 2. One-time step of the simulation model. Titles bounded by rectangles indicate either the points or the sides of the model overview (the triangle in Figure 1). The grids located on the green band show the territory of a parental species (in blue) and its daughters (red and green). The trees record their evolutionary relationship and time of speciation. For clarity "×" indicates the occurrence of local extinction when applying the influence of "Within Patch Dynamics" and "Between Patch Dynamics" to the territory grid. Here, unsuccessful long-distance dispersal and a speciation event occurred. All the parental territory was successfully divided up between the two daughter lineages.

sides) are: i. Within Cell Dynamics: models the effect of cell suitability on the local extinction rate to calculate the probability of species prevalence at time t within a cell. ii. Between Cell Dynamics models the colonisation success to a cell of a dispersing species through the relationship between cell suitability and dispersal success. iii. Evolutionary Events: models how the speciating species' range is divided up between its daughter lineages through the relationship between the speciation rate, the dispersal kernel, and the environmental characteristics of the cells (i.e. habitat suitability). Niche conservation, and constancy of speciation and background extinction rates are assumed to occur between and within species, but not necessarily between time steps.

MODEL ALGORITHM

In this section, we describe in more detail the model algorithm specifying the set of rules and functions that link the parameters in

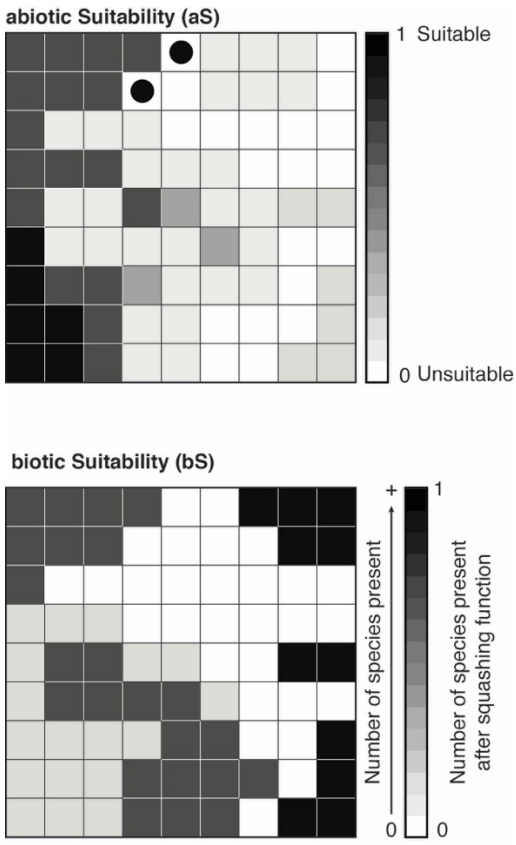


Figure 3. Abiotic and biotic suitability of each cell within the landscape. Suitability is composed of two components: *Abiotic suitability (aS)* probability is created by the environmental suitability (the climatic properties of the cell in relation to the tolerances of the species) and by physical formations (slow moving barriers). *Biotic Suitability (bS)* is the number of species present within a cell (density dependence effects). Abiotic and biotic suitability are bounded between $[0,1]$ values. Environmental abiotic suitability is calculated by an environmental niche model (ENM) that is hindcasted across the time steps t. Physical formations are slow moving barriers (mountains, rivers and deserts) that are that appear and disappear over time, and can move to a new geographical space (e.g. mountain building) Physical formations can be represented as cells of low suitability regardless as to whether the cell has high environmental suitability (black circle represents high environmental suitability, however due to a physical formation, the cell is considered to have low suitability).; for simplicity, we consider the landscape to be flat and so barriers are only barriers by width and not height. See more details in the text.

the model, as well as the distribution probability of these. Figure 2 shows an example of how the model algorithm works for one-time step t in which there is one event of long distance dispersal and one successful evolutionary event.

Habitat suitability

We considered the landscape suitability of a cell to be combination of two elements or values: a. abiotic suitability, which is dependent on the environmental suitability of the cell for the moving species and on the presence of temporary, slow moving barriers; b. biotic suitability, which is dependent on the number of species present within a cell (Figure 3).

Abiotic suitability (aS). This is a suitability probability generated by an environmental niche model (ENM) that has been hindcasted across the time steps of paleoclimate layers. There are a wide variety of models and distances that can be used in ENM hindcasting, including the Mahalanobis distance (Clark et al. 1998), the maximum entropy algorithm (Maxent; Phillips et al. 2006), or Hengl et al. (2009)'s framework combining distances defined by point-pattern analysis and environmental niche factor analysis (ENFA) with a regression-kriging GLM. Each of these models has their advantages and disadvantages (see Chefaoui and Lobo 2008). We use here an approach

suggested by Culshaw et al. (in rev.), where Hengl et al. (2009)'s method is used to generate an ENM hindcasted projection for each month of the year, and these monthly projections are then overlaid to construct a "yearly" projection. Culshaw et al. (in rev.) demonstrated that this approach is more efficient in capturing the general pattern than the usual approach of averaging the results across a collection of months (e.g., Mairal et al. 2017), as the accumulation of errors in the latter can result in the pattern being completely wiped out.

Slow moving barriers are considered to be physical formations, such as mountains, rivers and deserts, that appear and disappear over time, and, depending on the nature of the barrier, can move to a new geographical space, i.e., mountain building and erosion, or river capture, changes the position of a river basin. These physical formations can have a strong influence on the geographical distributions of a species, acting as barriers to or corridors for dispersal and migration, and, hence, deterring or facilitating movement between cells. These barriers can also be included in the aS matrix (i.e. values of aS per cell in the gridded domain) as low suitability areas. For simplicity, we consider the landscape to be flat, so that barriers are only composed of a width dimension and not of a height dimension. Moreover, we assume that the barrier can be overcome if the species is able to disperse past it to the next suitable cell in one-time step t, regardless of whether the terrain cells between a species and the colonised cell are highly unsuitable, hence, trivialising the role played by extreme ecological limitations (e.g. sharks are unable to walk on land and there is only land between them and another cell of water), see Connolly (2005) for the opposite model).

Once the model for estimating abiotic habitat suitability landscape for $t = 0$ for a specific area has been created, we interpolate this model back through time across all the paleoclimate data for the same area. The abiotic landscape suitability's values (aS) lie within the interval $[0, 1]$. However, as the aS matrix can be affected by loss of information caused by events such as population bottlenecking and mass extinction events (Culshaw et al. in rev, a Brownian motion distribution (Wiener process) is added or subtracted from each aS at each time step t.

Biotic Suitability (bS) refers to the density of the patch in regard to to the number of species. The bS matrix contains the number of species within each of the cells. These values are bound within the interval of $(0, 1]$ through the use of a squashing function $\overline{bS} = \frac{bS+a}{bS}$, where $a \in \mathbb{R}^{>0}$. Information on the number of species within each cell at any given time is relevant to know the level of interspecific competition present. This knowledge can be used to explain the absence of a given species from regions that are within their set of tolerable environmental conditions (Wiens and Donohue 2004).

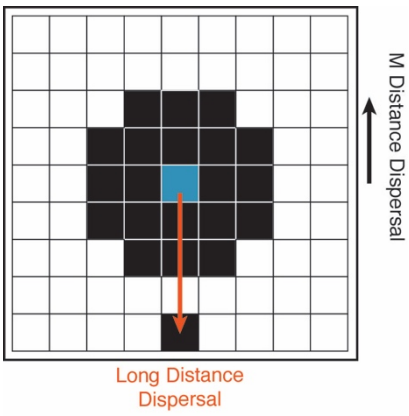


Figure 4. Range Evolution. Here individuals can disperse through the landscape by M distance in one-time step t or via a long distance and unlimited dispersal to any patch that is greater than M distance away. Long distance distribution is equally as likely of originating in any previously occupied cell.

Range evolution

Range evolution describes how individuals disperse through land (Figure 4). If the time step t is suitably large (e.g. 500 years), then, in the clear-cut thought process of an invasive species, the land should be completely colonised except for areas that cannot be reached by long distance dispersal, have patch unsuitability, or high competition (Rangel et al. 2007). Here, species in the assemblage share a common spreading dye distribution (Jetz and Rahbek 2001; Colwell et al. 2009; Gotelli et al. 2009), where an individual can move M distance in one-time step t but can also have long-distance unlimited dispersal to any patch more than M distance away with the probability distribution of this event occurring drawn from the Poisson distribution $FD = pois(\lambda_{FD})\, n_{FD}$, where magnitude $n_{FD} \in \mathbb{R}^{>0}$. Freak dispersal is equally likely of originating in any previously occupied cell.

Evolutionary Origins

During a simulation, this model allows for an evolutionary tree to be built in four different ways (Figure 5). In each of these tree types, the speciation rate (λ) and extinction rate (μ) must be supplied and the tree is built over a length of time that is given by the length of the simulation (T). Parameters λ and μ can vary over time, i.e., undergo rate shifts, but at any given time t rate homogeneity is assumed across lineages, even though this has been shown to be empirically unsound (Rabosky 2014). Since the model assumes phylogenetic niche conservation (i.e., the transmission from ancestor to descendants of the biological and physiological characteristics that determine the fundamental ecological niche of a species (Hutchinson 1957), the probability of a speciation event occurring at time step t is only determined by the speciation rate λ; this follows a Poisson distribution $pois(\lambda_\lambda)n_\lambda$, where $\lambda_\lambda \in \mathbb{R}^{>0}$ and magnitude $n_\lambda \in \mathbb{R}^{>0}$. The species that undergoes speciation is selected after all species in the simulation have dispersed for that time step t. Once a species has speciated, it is removed from the simulation. The rate of extinction, μ, is drawn from a Poisson distribution $pois(\lambda_\mu)n_\mu$, where $\lambda_\mu \in \mathbb{R}^{>0}$ and magnitude $n_\mu \in \mathbb{R}^{>0}$. In our model, the Poisson distribution of μ is reshaped by the

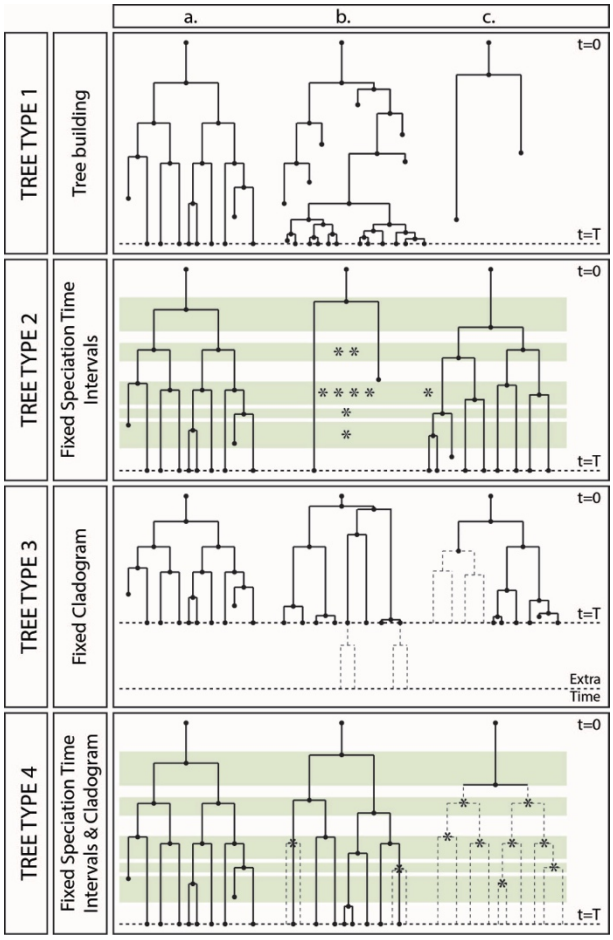


Figure 5. Evolutionary Origins. Evolutionary trees can be built in four different ways, which also represent constraints in the time and position of the speciation event. The figure presents three example trees that have been built under each of these four types for a maximum of $t = T$ time steps. *Tree type 1* allows simulations to produce any topology and any number of leaves or terminals. There is no cap on species number and there is no constraint in evolutionary relationships, as each species is as equally likely to speciate when it is time to speciate. *Tree type 2* constrains the time interval in which speciation may occur (here presented as a green band), but does not constrain the evolutionary relationships or tree topology, i.e. each species is equally likely to speciate at a given time. Each speciation time interval given covers at least one speciation opportunity (each marked with $*$). For tree "b", only one speciation event has occurred, with only one of the lineages reaching the present day. Trees "a" and "c" have a similar number of speciation events and the same number of extant, but differ in their tree topology. *Tree type 3* constrains the evolutionary relationships in the tree, i.e., the tree topology. This means that speciation is not constrained to occur within predefined time intervals, but when there is a speciation event, the species that undergoes speciation is predefined. In tree "b", due to lack of a successful speciation event, one daughter lineage has failed to manifest before the end of the simulation (the missing lineage is denoted by the dotted line); if the tree continues evolving past the simulation time, it may have a tree topology matching that of tree a ($t = T +$ extra time). In tree "c", daughters and granddaughters of a species have failed to speciate before the grandparent species became extinct (missing lineage shown by dotted line), and hence tree c can never fully fulfil its predefined topology. *Tree type 4* has the combined constraints of tree types 1 to 3: rates μ and λ, the time interval in which speciation is allowed to take place, and the tree topology. Unlike tree type "3b", tree type "4b" cannot be given extra time to follow its predefined tree topology (dotted line), as there are no more speciation time intervals during which to speciate. A more extended explanation is given in the text.

local patch dynamics occupied by the species (within patch dynamics), which determines the local extinction rate. Ultimately, the accumulation of local extinctions in different patches results in a species global extinction. In addition to these parameters, trees type 2, 3 and 4 require supplying additional information: the speciation time intervals in the case of tree type 2, information on the evolutionary relationships (i.e., the tree topology) for tree type 3; and both the speciation time intervals and information on the phylogenetic relationships for tree type 4. This information adds further constraints upon the building of the tree.

Tree type 1 allows simulations to produce any topology and any number of leaves or terminals. There is no cap on species number, and there are no constraints on evolutionary relationships, as each species is equally likely to speciate when it is time to

speciate. This tree building method follows the non-equilibrium model assumed by standard birth-death dynamics (Morlon 2014), defined only by μ and λ. It resembles the method for building evolutionary trees in GSMs suggested by Gotelli et al. (2009).

Tree type 2 constrains the time interval in which speciation may occur (represented in Figure 5 by a green band), but does not set any constraints on the tree topology, i.e. each species is equally likely to speciate when it is time to speciate. This tree building method is similar to the birth-death models that are conditional on a given initiation time (root age) and absolute speciation times, but where the number of terminal species can vary across reconstructed trees (Stadler 2011). Each time interval covers at least one speciation event (marked with $*$). In Figure 5, for example, two of the speciation time intervals contain several possible speciation events. This could be the case when speciation time intervals overlap or are so close that they can be considered to be the same event. It is at the user's discretion to decide whether single or multiple speciation events are allowed within a single time interval. A speciation event is not required to occur, only that if it were to occur, it can only occur within the speciation time intervals. Therefore, the absolute value for time to speciation (branch length) is only partially predefined. For tree "b" only one speciation event has occurred, with only one of the descendant lineages reaching the present day. This implies that speciation failed to occur for each of the daughter lineages within a speciation time interval before either the descendant became extinct or the simulation ended. Trees "a" and "c" have a similar number of speciation events (9 and 8 respectively) and the same number of extant species in the tree; however, they differ in their topologies (Figure 5).

Tree type 3 constrains the evolutionary relationships in the tree, i.e., the tree topology. This means that speciation is not constrained to occur within predefined time intervals, but when there is a speciation event, the species that undergoes speciation is predefined. As in tree type 2, a species is not dictated to speciate; hence, the absolute value for time to speciation (branch length) is not considered. However, should a species fail to speciate, then all its possible daughters will not speciate either, i.e. a descendant speciation event cannot precede that of its ancestor. In Figure 5, tree "b", due to lack of a successful speciation event, one daughter lineage has failed to manifest before the end of the simulation (the missing lineage is denoted by a dotted line). In theory, should tree "b" be given extra simulation time, it may exhibit a tree topology matching that of tree a. In tree "c", daughters and granddaughter lineages of one of the initial species have failed to speciate before the grandparent species became extinct (missing lineage shown by dotted line), and hence, tree "c" can never fully fulfil its predefined tree topology (Figure 5).

Tree type 4 has the combined constraints of tree types 1 to 3: rates μ and λ, the time interval in which speciation is allowed to take place, and the tree topology. Thus, as in tree type 2, the absolute value for time to speciation (branch length) is predefined. Unlike tree type "3b", tree type "4b" cannot be given extra time to follow its predefined tree topology (dotted line), as there are no more speciation time intervals during which to speciate (Figure 5).

Having these different tree types is advantageous. If the user has more information than just λ and μ, he can make a more informative simulation including lineage divergence times or tree topology, and this allows for a more direct comparison with empirical or previously simulated phylogenies. On the other hand, if the user has doubts or lacks estimations of μ and/or λ, then, information about only the topology can be used to explore the μ and λ parameter space to find a combination of these two rates that works. This is particularly interesting because estimations of μ are in general considered to be problematic to obtain (Sanmartin and Meseguer 2016), while estimates of λ and speciation times are deemed reliable.

Extinction

Extinction (local, eventually resulting in global) and successful colonisation can be important explanations for the absence of a clade from a given area. Although extinction needs not directly involve dispersal, the absence of a clade from a region owing to extinction or unsuccessful (re)colonisation indicates a limitation on dispersal (Wiens and Donohue 2004). In our model, extinction and colonisation success following dispersal are not equiprobable among grid cells; instead, they are dependent on the particular values the environmental variables adopt in each cell.

Within and between patch dynamics

Within and between patch dynamics models the relationship between habitat suitability and extinction rate (μ), and suitability and the probability of colonisation success ($\mathbb{P}(C)$) (Figure 6). $\mathbb{P}(C)$ may undergo rate shifts through the simulation. It is known that the probability of extinction is dependent on abiotic or biotic factors (i.e., passive replacement and active displacement, Silvestro et al. 2015). This refers to the local extinction of individuals within species; if all individuals of a species are removed, then a global extinction occurs and the species can no longer speciate and leave behind daughter lineages. In our model, we combine the abiotic suitability matrix (aS) with the biotic suitability matrix (bS) (Figure 6a) in order to calculate the magnitude of the influence of the local environmental conditions in the patch (i, j) at time step t over the Poisson distribution μ and $\mathbb{P}(C)$ are drawn from, separately; this allows

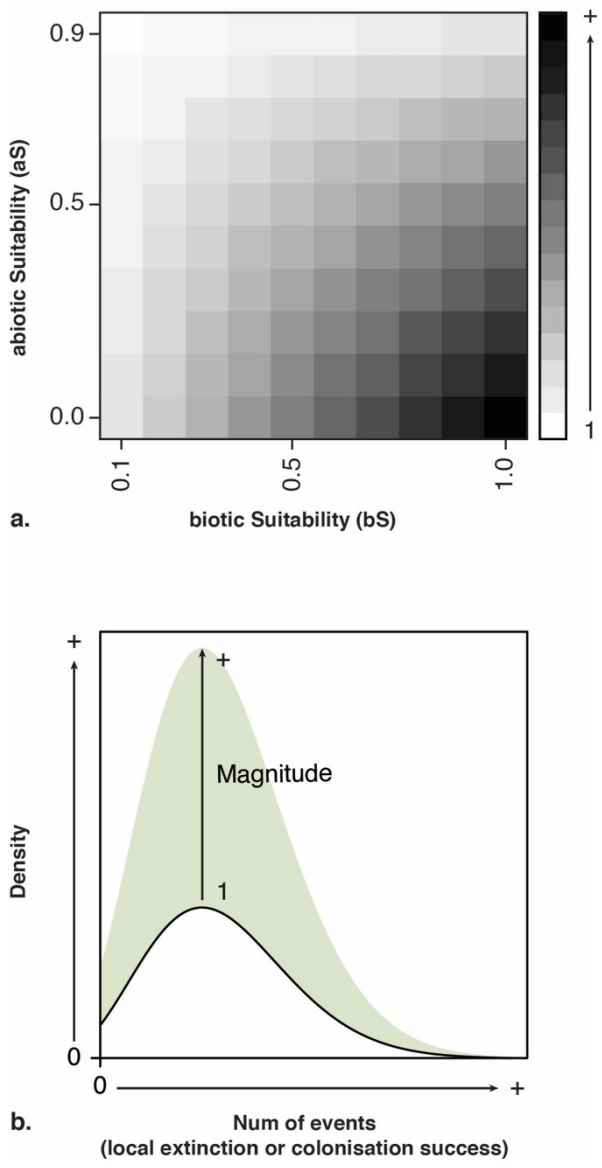


Figure 6. Demonstrating how local background extinction rate is calculated from values of cell values of abiotic and biotic values at time t. a. The model default to combine aS and $\overline{bS}$ together is linear $(1 - aS)\overline{bS}$, where $\overline{bS}$ = shaper function(bS) member $(0, 1]$ as there is always at least one species present in a patch that local background extinction is calculated for and $aS \neq 1$ as conditions can never be 100% perfect. The default magnitude is exponential and its gradient is controlled by $n \in \mathbb{R}^{>0}$. As $n \to 0$, the influence of the magnitude over the μ Poisson distribution becomes minimal to non existent, e.g. each patch has enough resources for all. a. shows the default magnitude with any n. Notice that as $aS \to 1$ and $bS \to 0$ then the magnitude $\to 1$ i.e. the μ Poisson distribution and colonisation success remain unchanged.

calculation of the local extinction rate and colonisation success (Figure 6b)

The default model to combine aS and $\overline{bS}$ together is linear $(1 - aS)\overline{bS}$, where $\overline{bS}$ = shaper function(bS) member $(0, 1]$, as there is always at least one species present in a patch for which local extinction is calculated, and $aS \neq 1$ as conditions can never be 100% perfect. The default magnitude is exponential and its gradient is controlled by $n \in \mathbb{R}^{>0}$. As $n \to 0$, the influence of the magnitude over the μ Poisson distribution becomes minimal to non-existent, e.g. each patch has enough resources for all species. Figure 6a shows the default magnitude with any n. As $aS \to 1$ and $bS \to 0$, the magnitude $\to 1$ i.e. the μ Poisson distribution and colonisation success remain unchanged. Figure 6b presents the Poisson distribution for local extinction μ at patch (i, j) at time t:

$$pois\big(local\ \lambda_{\mu\,(i,j,t)}\big) = pois\big(\lambda_{\mu}\big)\, n_{\mu}\, e^{\left(\left(1-aS_{(i,j,t)}\right) bS_{(i,j,t)}\right)^{n}}$$

The probability for local colonisation success at patch (i, j) at time t is:

$$\mathbb{P}\big(local\ C_{(i,j,t)}\big) = \mathbb{P}(C)\, e^{\left(\left(1-aS_{(i,j,t)}\right) bS_{(i,j,t)}\right)^{-n}}$$

The user can change all default models.

Division of parental territory

An Evolutionary Event occurs when Lineage Information tells the model that a

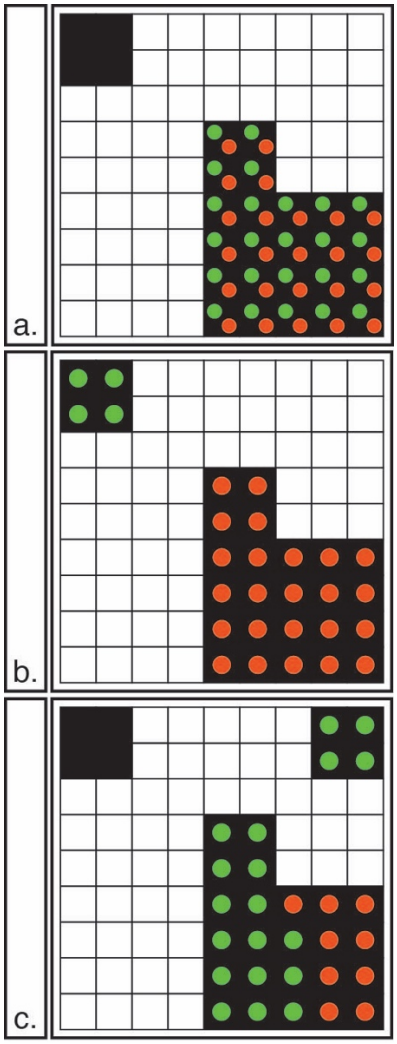


Figure 7. Range inheritance scenarios: The figure shows the division of the parental territory between the daughters at the time of an evolutionary event. The parent's territory is denoted in black, with each daughter denoted as a red or green dot. There are three ways in which parental territory can be assigned to the daughters. a. Daughters originate from the same cell. Both daughters will inherit the same territory. b. The cells lie M distance away from only one origin patch. The territory is divided up without conflict. c. Daughters originate from different cells but have the possibility to inherit the same territory. In this case the territory will be divided up, with patches *M* distance away from both origin cells being assigned to the daughter that is closer. If the daughters are the same distance away from the cell being assigned, the conflict is resolved by pseudo randomly assigning it to a daughter.

speciation event has occurred, once that species has dispersed (Figure 2). In our model, the Evolutionary Event component of the model addresses the geographical mode of speciation by assigning origin patches, chosen from the parent's territory, to the daughters (in Figure 7, these are denoted in black, and with red and green dots, respectively). All patches that form the parental territory are equally likely to be selected. After this, the parent's territory is divided between the two daughter lineages. There are three ways a parental territory can be divided into its daughters. Notice that only territory that is *M* distance away from the origin patch is considered; any unreachable territory is disregarded, i.e. black patches lacking a green or red dot in Figure 7.

a. Daughters originate from the same patch; both daughters will inherit the same territory.
b. Daughters originate from different patches and do not have the possibility to inherit the same territory. The patches lie M distance away from only one origin patch. The territory is divided up without conflict.
c. Daughters originate from different patches but have the possibility to inherit the same territory. In this case, the territory will be divided up, with patches M distance away from both origin patches being assigned to the daughter that is closer. If the daughters are the

same distance away from the patch being assigned, the conflict is resolved by pseudo randomly assigning it to a daughter.

Once a parent species has speciated, it is removed from the simulation. The three models of fragmentation of the parental territory represent the sympatric, parapatric or peratric/allopatric speciation modes (Gotelli et al. 2009).

CONCLUSIONS AND SOME SUGGESTIONS FOR TESTING THE MODEL

Rangel et al. (2007) evaluated the robustness of their GSM model in terms of its ability to predict long-discussed tenets in biogeography, such as the latitudinal diversity gradient or the mid-domain effect based on the distribution of 3000 species of South American birds. Predicted spatial patterns closely resemble observed ones and proved sensitive to niche dynamics processes. This allowed them to evaluate the roles of both evolutionary and ecological processes in explaining spatial patterns in species richness and geographic range size distribution, thus providing a "link between ecology and historical biogeography under integrated theoretical and methodological frameworks". Their GSM model was novel in using (artificially) fluctuating environmental conditions and a single ancestral origin for the species generated in the model, thus providing a link between speciation events.

In our general simulation model, we provide a link between speciation events in different way by instead of focusing on species richness patterns or the predicted spatial distribution of species (a pattern-oriented modelling approach (Grimm et al. 1996, 2005), we propose to test our model by explicitly introducing the evolutionary component in the form of a phylogeny and associated lineage divergence times. We select parameter values for the final simulation based on maximising the similarity between observed and simulated spatial patterns, specifically: i. the geographic distribution of extant species; and ii. the resultant phylogenetic pattern from the simulations. Accounting for phylogenetic relationships and a time dimension in comparing the simulated and the real world is actually important to increase the realism of the model and its power to detect system-level properties (Gotelli et al. 2009). The discipline of biogeography has both spatial and temporal dimensions, and the level of niche conservatism can change over the history of a lineage (Sanmartín 2014). The shape and length of a phylogeny can also be informative about the actual processes involved in their generation in the first place: for example, "broom-and-handle shape" phylogenies are typically associated with high background extinction rates or mass extinction events (Sanmartín and Meseguer 2016). Hence, introducing phylogenetic shape

and time in the simulation outcomes might help us test the role played by background extinction and high extinction rates in the formation of both spatial patterns and evolutionary patterns.

Other differences between our model and Rangel et al. (2007) lies in the modelling of extinction and the introduction of actual empirical data in the model. In their GSM model, speciation is caused by fragmentation and isolation, while extinction is linked to reductions in range size up to a certain threshold. In our model, speciation is brought about via a speciation rate, and extinction by an extinction rate that is dependent on abiotic (palaeoenvironmental (climate)) and biotic (density-dependence) factors, thus helping introduce a basic species competition to the model. Also in our GSM, the speciation event can be further informed with speciation time intervals and the empirical tree cladogram. This allows for a more direct comparison of the GSM results to an empirical example, but also interestingly, this lends to an opportunity to explore the speciation and extinction rate space to find suitable pairings of speciation and extinction rates for phylogenies that are unsuitable for traditional birth-death model. Rather than using a cyclical environmentally fluctuating landscape as in Rangel and collaborators, we use a time series of palaeoclimatic layers spanning the last 20 million years. Using empirical data, our understanding of the evolutionary relationships and biogeographic history of real organisms (Gotelli et al. 2009) can help better refine and constrain the range of parameter values examined during the simulations. This might be especially important when we do not have a large phylogenetic and distribution dataset as input.

One nice example of such a case comes from a biogeographic pattern of plant distributions in Africa tackled by our team. The Rand Flora (RF) pattern is a continental-scale geographic disjunction in which sister species are distributed on opposite sides of the African continent: in Macaronesia-northwestern Africa, Eastern Africa-southern Arabia and southern Africa (Sanmartín et al. 2010). Dated phylogenies and biogeographic analyses in some of these groups suggest that most RF disjunctions date from the Early Miocene to the Mid Pliocene periods, concurrent with an aridification trend in Africa, a result of plate tectonics and mountain building (Pokorny et al. 2015). Species populations are often small in size and have highly restricted distributions. Recent ENM reconstructions (Mairal et al. 2017b) suggest that RF-lineage distributions were broader in the past and that range expansion was achieved by climatic corridors, which were later interrupted by successive aridity waves. One difficulty with inferring the origin of the RF biogeographic pattern is that most RF phylogenies comprise of less than 5-10 species and often exhibit “broom-and-handle” shapes, with long stem-branches and young crown clades, indicative of high

extinction rates (Sanmartín and Meseguer 2016). The long time intervals without cladogenetic events along the phylogeny hinders the use of birth-death macroevolutionary and biogeographic inferential models. Culshaw et al. (in rev.) handled this uncertainty by adopting an integrative inferential approach, combining information from biogeographic analyses and ENMs to disentangle the processes behind the RF pattern in the small (only three species) but highly disjuct (continental distribution) genus *Camptoloma* (Scrophulariaceae). As in many RF lineages, information on species and population distributions is scarce.

The spatially-explicit forward time computational simulation model could be also another avenue to tackle the RF pattern. For example, one could use our GSM model and paleoclimate data from Africa (as done in Culshaw et al. in rev. and Mairal et al. (2017) to try and recreate the Rand Flora distribution pattern (continental disjunction between the margins of Africa) and the evolutionary phylogenetic shapes characterising RF lineages (ancient stem-age, young crown-group and long internodes). One advantage of using the GSM approach is that we can let estimates of diversification and background extinction rates obtained from the birth-death framework (Pokorny et al. 2015) to inform or constrain the range of values adopted by speciation and extinction rates in the model), thus increasing the realism of the model in relation to the examined pattern, but also helping to reduce the uncertainty in the comparison between observed and simulated patterns. One strong assumption of our model is that we assume niche conservatism over the entire simulation. Species expand their range occupying adjacent cells with similar conditions. Rangel et al. (2007)'s GSM allowed niche evolution, by modelling random small deviations, from the niche breadth centre using a Brownian model (sigma). Assuming niche conservatism is risky for long geological time scales. For the Rand Flora, this is not such a crucial impediment because biogeographic inferences seem to point out to a conservation of ancestral tolerances and fragmentation being responsible for the RF distribution (Rincón-Barrado et al. in rev.; Culshaw et al. in rev.); it is also a more conservative approach in the absence of information. Moreover, Rangel et al. (2007)'s model, initial tolerance to climatic conditions is determined by the initial location of the first simulated species (i.e., climatic conditions of the starting cell). In our model, this would be given by the actual ENM models based on empirical data and by the real palaeoclimate layers for our specific biogeographic domain (Africa), thus again reducing uncertainty in the model.

In sum, we argue that the computational general simulation model presented here could be an interesting alternative to explore hypotheses on the spatial patterns and processes underlying general distribution

biodiversity patterns, especially when the initial input data is small, thereby introduces large uncertainties in the model.

Acknowledgments: We thank Andrew Hendry for the helping with the initial concept of the model, Joaquín Hortal for providing the introduction to Thiago Rangel. Thiago Rangel for welcoming VC into LETS, Universidade Federal de Goiás, Brasil. VC was funded by Ministerio de Economía y Competitividad MINECO Ph.D. (FPI) Fellowship BES-2013-065389, supervised by IS. IS was funded by the Spanish Government and European Regional Development Fund, grant CGL2015-67849-P (MINECO/FEDER).

Author contributions: VC designed the model under the guidance of TFLVBR and help from IS. VC and IS co-wrote the manuscript.

REFERENCES

Antonelli A., Sanmartín, I. 2011. Mass extinction, gradual cooling, or rapid radiation: Reconstructing the spatiotemporal evolution of the ancient angiosperm genus *Hedyosmum* (Chloranthaceae) using empirical and simulated approaches. Systematic Biology. 60:596-615.

Araújo M.B., Pearson R.G., Thuiller W., Erhard M. 2005 Validation of species–climate impact models under climate change. Global Change Biology. 11:1504-1513.

Araújo M.A., Luoto M. 2007. The importance of biotic interactions for modelling species distributions under climate change. Global Ecology and Biogeography. 16:743-753.

Araújo M.B., Ferri-Yáñez F., Bozinovic F., Marquet P.A. Valladares F., Chown S.L. 2013 Heat freezes niche evolution. Ecology Letters. 16:1206-1219.

Basille M., Calenge C., Marboutin E., Andersen R., Gaillard, J.M. 2008. Assessing habitat selection using multivariate statistics: Some refinements of the ecological-niche factor analysis. Ecological Modelling. 211: 233-240.

Beerling D.J., Berner R.A., Mackenzie F.T., Harfoot M.B., Pyle J.A. 2009. Methane and the CH4 related greenhouse effect over the past 400 million years. American Journal of Science. 309:97-113.

Beerling D.J., Fox A., Stevenson D.S. Valdes P.J. 2011 Enhanced chemistry-climate feedbacks in past greenhouse worlds. Proceedings of the National Academy of Science USA. 108:9770-9775.

Beerling D.J., Taylor L.L., Bradshaw C.D.C., Lunt D.J., Valdes P.J., Banwart S.A., Pagani M., and Leake J.R. 2012. Ecosystem CO_2 starvation and terrestrial silicate weathering: mechanisms and global-scale quantification during the late Miocene. Journal of Ecology. 100: 31-41.

Boone R.B. 2010. Simulating species richness using agents with evolving niches, with an example of Galápagos plants. International Journal of Ecology. 2010:doi: 10.1155/2010/150606.

Bradshaw C.D., Lunt D.J., Flecker R., Salzmann U., Pound M.J., Haywood A.M., Eronen J.T. 2012. The relative roles of CO_2 and palaeogeography in determining Late Miocene climate: Results from a terrestrial model-data comparison. Climate of the Past Discussions. 8:1257-1258.

Brooks D.R. 1990. Parsimony analysis in historical biogeography and coevolution: methodological and theoretical update. Systematic Zoology. 39:14-30.

Brooks D.R., McLennan D.A. 1991. Phylogeny, ecology, and behavior: a research program in comparative biology. Chicago: University of Chicago Press.

Chefaoui R.M., Lobo J.M. 2008. Assessing the effects of pseudo-absences on predictive distribution model performance. Ecological Modelling. 210:478-486.

Clark J.D., Hayes S.G., Pledger J.M. 1998. A female black bear denning habitat model using a geographic information system. Ursus. 10:181-185.

Colwell R.K., Gotelli N.J., Rahbek C., Entsminger G.L., Farrell C., Graves G.R. 2009. Peaks, plateaus, canyons, and craters: the complex geometry of simple mid-domain effect models. Evolutionary Ecology Research. 11:355-370.

Connolly S.R. 2005. Process-based models of species distributions and the mod-domain effect. The American Naturalist. 166:1-11.

Cox, Moore and Ladle. Biogeography: An Ecological and Evolutionary Approach, 10th edition, 2016.

Cracraft J. 1994. Species-diversity, biogeography, and the evolution of biotas. American Zoologist. 34:33-47.

Craven P., Craven D. 2000. The flora of the Brandberg, Namibia. Cimbebasia Memoir. 9: 49-67.

Culshaw V., Mairal M., Sanmartín I. rev. Combing Bayesian biogeographic inference and phylogenetically-informed niche models to reconstruct the role of ancient climate change in depauperate lineages. Systematic Biology.

Currie D.J., Francis A.P., Kerr J.T. 1999. Some general propositions about the study of spatial patterns of species richness. Ecoscience. 6:392-399.

De Marco Jr P., Diniz-Filho J.A.F., Bini L.M. 2008. Spatial analysis improves species distribution modelling during range expansion. Biology Letters. 4:577-580.

Domina G., Marino P., Spadaro V., Raimondo F.M. 2012. Vascular flora evolution in the major Mediterranean islands. Biodiversity Journal. 3:337-342.

Dormann C.F., McPherson J.M., Araújo M.B., Bivand R., Bolliger J., Carl G., Davies R.G., Hirzel A., Jetz W., Kissling W.D., Kühn I., Ohlemüller R., Peres-Neto P.R., Reineking B., Schröder B., Schurr F.M., Wilson R. 2007. Methods to account for spatial autocorrelation in the analysis of distributional species data: a review. Ecography. 30:609-628.

Evans M.E.K., Smith S.A., Flynn R.S., Donoghue M.J. 2009. Climate, niche evolution, and diversification of the "bird-cage" evening primroses (*Oenothera*, Sections *Anogra* and *Kleinia*). The Ameriacn Naturalist. 173:225-240.

Fischer A.G. 1960. Latitudinal Variations in Organic Diversity. Evolution. 14:64-81.

Gotelli N.J., Anderson M.J., Arita H.T., Chao A., Colwell R.K., Connolly S.R., Currie D.J., Dunn R.R., Graves G.R., Green J.L., Grytnes J.-A., Jiang Y.-H., Jetz W., Lyons S.K., McCain C.M., Magurran A.E., Rahbek C., Rangel T.F.L.V.B., Soberón J., Webb C.O., Willig M.R. 2009. Patterns and causes of species richness: a general simulation model for macroecology. Ecology Letters. 12:873-886.

Grimm V., Frank K., Jeltsch F., Brandl R., Uchmański J., Wissel C. 1996. Pattern-oriented modelling in population ecology. The Science of the Total Environment 183:151-166.

Grimm V., Revilla E., Berger U., Jeltsch F., Mooij W.M., Railsback S.F., Thulka H.-H., Weiner J., Wiegand T., DeAngelis DL. 2005. Pattern-oriented modeling of agent-based complex systems: lessons from ecology. Science 310:987-991.

Grytnes J.A. 2003 Ecological interpretation of the mid-domain effect. Ecology Letters 6:883-888.

Guisan A., Rahbek C. 2011. SESAM–a new framework integrating macroecological and species distribution models for predicting spatio- temporal patterns of species assemblages. Journal of Biogeography. 38:1433-1444.

Hassell M.P., Comins H.N., May R.M. 1991. Spatial structure and chaos in insect population dynamics. Letters to Nature. 353:255-158.

Hawkins B.A., Field R., Cornell H.V., Currie D.J., Guegan J.F., Kaufman D.M., Kerr J.T., Mittelbach G.G., Oberdorff T., O'Brien E.M., Porter E.E., Turner J.R.G. 2003. Energy, water, and broad-scale geographic patterns of species richness. Ecology. 84:3105-3117.

Hawkins B.A., Diniz-Filho J.A.F., Weis A.E. 2005 The mid-domain effect and diversity gradients: is there anything to learn. The American Naturalist 166:140-143.

Hengl T., Sierdsema H., Radović A., Dilo A. 2009. Spatial prediction of species' distributions from occurrence-only records: combining point pattern analysis, ENFA and regression-kriging. Ecological Modelling. 220:3499-2511.

Hillebrand H. 2004 On the generality of the latitudinal diversity gradient. The American Naturalist. 163:192-211.

Hirzel A.H., Hausser J., Chessel D., Perrin N. 2002 Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? Ecology 83:2027-2036.

Hubbell S.P. 2001. The unified neutral theory of biodiversity and biogeography. Princeton University Press.

Hutchinson G.E. 1957. Population studies —animal ecology and demography— concluding remarks. Cold Spring Harbor Symposia on Quantitative Biology. 22:415-427.

Jablonski D. 2008. Extinction and spatial dynamics of biodiversity. Proceedings of the National Academy of Science USA. 105:11528-11535.

Jetz W., Rabnek C. 2001 Geometric constraints explain much of the species richness pattern in African birds. National Academy of Sciences 98:5661-5666.

Judson O.P. 1994 The rise of the individual-based model in ecology. Proceedings of the National Academy of Science USA. 98:5661-5666.

Kreft H., Jetz W. 2007. Global patterns and determinants of vascular plant diversity. Proceedings of the National Academy of Sciences USA. 104:5925-5930.

Lemey P., Rambaut A., Drummond A.J., Suchard M.A. 2009. Bayesian phylogeography finds its roots. PLoS Computational Biology. 5:e1000520. doi:10.1371.

Levins R. 1969. Some demographic and genetic consequences of environmental heterogeneity for biological control. Bulletin of the Entomological Society of America 15:237-240.

Lobo J.M., Jiménez-Valverde A., Real R. 2008. AUC: a misleading measure of the performance of predictive distribution models. Global Ecology and Biogeography. 17:145-151.

MacArthur R., Wilson E.O. 1967. The theory of biogeography. Princeton University Press, New

Jersey:19-67.

Mairal M., Pokorny L., Aldasoro J.J., Alarcón M., Sanmartín I. 2015. Ancient vicariance and climate-driven extinction explain continental-wide disjunctions in Africa: the case of the Rand Flora genus *Canarina* (Campanulaceae). Molecular Evolution. 24:1335-1354.

Mairal M., Sanmartín I., Pellissier L. 2017. Lineage-specific climatic niche drives the tempo of vicariance in the Rand Flora. Journal of Biogeography. 44:doi: 10.1111/jbi.12930.

Mairal M., Caujapé-Castells J., Pellissier L., Jaén-Molina R., Álvarez N., Heuertz M., Sanmartín I. 2018. A tale of two forests: ongoing aridification drives population decline and genetic diversity loss at continental scale in Afro-Macaronesian evergreen-forest archipelago endemics. Annals of botany. 122:1005-1017.

Meseguer A.S., Lobo J.M., Ree R., Beerline D.J., Sanmartín I. 2015. Integrating fossils, phylogenies, and niche models into biogeography to reveal ancient evolutionary history: The case of *Hypericum* (Hypericaceae). Systematic Biology. 64:215–232.

Meseguer A.S., Lobo J.M., Cornuault J., Beerling D., Ruhfel B.R., Davis C.C., Jousselin E., Sanmartín I. 2018. Reconstructing deep-time palaeoclimate legacies in the clusioid Malpighiales unveils their role in the evolution and extinction of the boreotropical flora. Global Ecology and Biogeography. 27:616-628.

Mittelback G.G., Schemske D.W., Cornell H.V., Allen A.P., Brown J.M., Bush M.B., Harrison S.P., Hurlbert A.H., Knowlton N., Lessios H.A., McCain C.M., McCune A.R., McDade, L.A., McPeek M.A., Near T.J., Price T.D., Ricklefs R.E., Roy K., Sax D.F., Schluter D., Sobel J.M., Turelli M. 2007. Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. Ecology Letters. 10:315-331.

Morlon H. 2014. Phylogenetic approaches for diversification. Ecology Letters 17:508-525.

Otto-Bliesner B.L., Marshall S.J., Overpeck J.T., Miller G.H., Hu A., CAPE Last Interglacial Project members 2006. Simulating arctic climate warmth and icefield retreat in the last interglaciation. Science 311:1751-1753.

Owens H.L., Campbell L.P., Dornak L.L. Saupe E.E., Barve N., Soberón J., Ingenloff K., Lira-Noriega A., Hensz C.M., Myers C.E., Peterson A.T. 2011. Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas. Ecological Modelling. 263:10-18.

Oxelman B., Kornhall P., Olmstead R.G., Bremer B. 2005. Further disintegration of Scrophulariaceae. TAXON. 54:411-425.

Pearson R.G., Thuiller W., Araújo M.B., Martinez-Meyer E., Brostons L., McClean C., Miles L., Seurado P., Dawson T.P., Lees D.C. 2006. Model-based uncertainty in species range prediction. Journal of Biogeography. 33:1704-1711.

Pellissier L., Bråthen K.A., Vittoz P., Yoccoz N.G., Dubuis A., Meier E.S., Zimmermann N.E., Randin C.F., Thuiller W., Garraud L., Van Es J., Guisan A. 2013. Thermal niches are more conserved at cold than warm limits in arctic-alpine plant species. Global Ecology and Biogeography 22:933-941.

Phillips S.J., Anderson R.P., Schapire R.E. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling. 190:231-259.

Pianka E.R. 1966 Latitudinal gradients in species diversity: a review of concepts. The American Naturalist. 100:65-75.

Pokorny L., Riina R., Mairal M., Meseguer A.S., Culshaw V., Cendoya J., Serrano M., Carbajal, R., Ortiz S., Heuertz M., Sanmartín I. 2015. Living of the edge: timing of Rand Flora disjunctions congruent with ongoing aridification in Africa. Frontiers in Genetics. 6:doi: 10.3389/fgene.2015.00154.

Pulliam H.R. 2000. On the relationship between niche and distribution. Ecology Letters. 3:349-361.

Qian H., Ricklefs R.E. 2004. Taxon richness and climate in angiosperms: is there a globally consistent relationship that

precludes region effects? The American Naturalist. 163:773-779.

Rabosky D.L. 2014. Automatic Detection of Key Innovations, Rate Shifts, and Diversity-Dependence on Phylogenetic Trees. PLOS ONE. 9:e89543.

Rahbek C., Gotelli N.J., Colwell R.K., Entsminger G.L., Rangel T.F.L.V.B., Graves G.R. 2007. Predicting continental-scale patterns of bird species richness with spatially explicit models. Proceedings of the Royal Society B: Biological Sciences. 274:165-174.

Rangel T.F.L.V.B., Diniz-Filho J.A.F., Bini L.M. 2006. Towards an integrated computational tool for spatial analysis in macroecology and biogeography. Global Ecology and Biogeography. 15:321-327.

Rangel T.F.L.V.B., Diniz-Filho J.A.F., Colwell R.K. 2007 Richness and evolutionary niche dynamics: a spatial pattern-oriented simulation experiment. American Society of Naturalists. 170:602-616.

Ree R.H., Smith S.A. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. Systematic Biology. 57:4-14.

Richards C.L., Carstens B.C., Knowles L.L. 2007. Distribution modelling and statistical phylogeography: an integrative framework for generating and testing alternative biogeographical hypotheses. Journal of Biogeography. 34:1833-1845.

Ricklefs, R. E. 2006a Time, species, and the generation of trait variance in clades. Systematic Biology. 55:151-159.

Ricklefs, R. E. 2006b Global variation in the diversification rate of passerine birds. Ecology. 87:2468-2478.

Ronquist F., Sanmartín, I. 2011. Phylogenetic methods in biogeography. Annual Review of Ecology, Evolution, and Systematics. 42:441-464.

Sanmartín I., Engho H., Ronquist F. 2001. Patterns of animal dispersal, vicariance and diversification in the Holarctic. Biological Journal of the Linnean Society. 73:345-390.

Sanmartín I., van Der Mark P., Ronquist F. 2008. Inferring dispersal: A Bayesian approach to phylogeny-based island biogeography, with special reference to the Canary Islands Journal of Biogeography. 35:428-449.

Sanmartín I., Anderson C.L., Alarcon M., Ronquist F., Aldasoro J.J. 2010. Bayesian island biogeography in a continental setting: The Rand Flora case. Biology Letters. 6:703-707.

Sanmartín I. 2014. Biogeography. Pp 555-575 in Vargas. P., Zardoya R. ed. The tree of life. Sinauer, Sunderland, Massachusetts, USA.

Sanmartín I., Meseguer A.S. 2016. Extinction in phylogenetics and biogeography: from timetrees to patterns of biotic assemblage. Frontiers in Genetics. 7:35.

Saupe E.E., Barve N., Owens H.L., Cooper J.C., Hosner P.A., Peterson A.T. 2017. Reconstructing ecological niche evolution when niches are incompletely characterized. Systematic Biology. 67:428-438.

Silvestro D., Antonelli A., Salamin N., Quental T.B. 2015. The role of clade competition in the diversification of North American canids. Proceedings of the National Academy of Science USA. 112:8684-8689.

Stadler T. 2011. Mammalian phylogeny reveals recent diversification rate shifts. Proceedings of the National Academy of Science USA. 108:6187-6192

Stevens G.C. 1989. The latitudinal gradient in geographical range: how so many species coexist in

the tropics. The American Naturalist. 133:240-256.

Storch D., Davis R.G., Zajíček S., Orme C.D.L., Olson V., Thomas G.H., Ding T.-S., Rasmussen R.C., Ridgely R.S., Bennett, P.M., Blackburn T.M., Owens I.P.F., Gaston K.J. 2006. Energy, range dynamics and global species richness patterns: reconciling mid-domain effects and environmental determinants of avian diversity. Ecology Letters. 9:1308-1320.

Thiv M., Thulin M., Hjertson M., Kropf M., Linder H.P. 2010. Evidence for a vicariant origin of Macaronesian–Eritreo/Arabian disjunctions in *Campylanthus* Roth (Plantaginaceae). Molecular Phylogenetics and Evolution. 54:607-616.

Thuiller W. 2003. BIOMOD–optimizing predictions of species distributions and projecting potential future shifts under global change. Global change biology. 9:1353-1362.

Thuiller W., Lavorel S., Sykes M.T., Araújo M.B. 2006. Using niche-based modelling to assess the impact of climate change on tree functional diversity in Europe. Diversity and Distributions. 12:49-60.

Thuiller W., Albert C., Araújo M.B., Berry P.M., Cabeza M., Guisan A., Hicker T., Midgley G.F., Paterson J., Schurr F.M., Sykes M.T., Zimmermann N.E. 2008. Prediction global change impacts on plant species' distributions: Future challenges. Perspectives in Plant Ecology, Evolution and Systematics. 9:137-152.

Villaverde T., Pokorny L., Olsson S., Rincón-Barrado M., Johnson M.G., Gardner E.M., Wickett N.J., Molero J., Riina R., Sanmartín I. 2018. Bridging the micro- and macroevolutionary levels in phylogenomics: Hyb-Seq solves relationships from populations to species and above. New Phytologist. 220:636-650.

Wiens J.J., Donoghue M.J. 2004. Historical biogeography, ecology and species richness. TRENDS in Ecology and Evolution. 19:639-644.

Wiens J.J. 2004. Speciation and ecology revisited: phylogenetic niche conservatism and the origin of species. Evolution. 58:193–197.

Wiley E.O. 1981. Phylogenetics: The Theory and Practice of Phylogenetic Systematics. Wiley-Interscience.

Willig M.R., Kaufman D.M., Stevens R.D. 2003. Latitudinal gradients of biodiversity: pattern, process, scale, and synthesis. Annual Review of Ecology, Evolution, and Systematics. 34:273-309.

CONCLUSIONES GENERALES EN CASTELLANO

CONCLUSIONES GENERALES EN CASTELLANO

El principal objetivo de esta Tesis Doctoral es desarrollar nuevas herramientas y metodologías para comprender la señal las respuestas evolutiva y biogeográficas al cambio climático en linajes de plantas. Para ello se emplearon cuatro enfoques distintos: modelos macroevolutivos de diversificación, métodos de inferencia estadística biogeográfica, modelización de nicho proyectada en escenarios pasados y futuros, y modelos de simulación computacional espacialmente explícitos. Estos enfoques se aplicaron a dos casos de estudio: la evolución del grupo de las coníferas (Gimnosperma, Coniferae; Capítulo 1), con distribución cosmopolita, y el género africano *Camptoloma*, con una distribución continental disyunta, la Rand Flora (Capítulos 2 y 3).

1. Se propone un nuevo modelo bayesiano de macroevolución (BDSKY) para explorar la señal evolutiva de la extinción masiva – eventos evolutivos que extinguen al menos el 75% de la diversidad existente en un corto período de tiempo geológico. Este modelo expande modelos episódicos *(skyline birth-death model*) utilizados en el campo de la epidemiología, a filogenias moleculares de especies que abarcan millones de años y que carecen de información del registro fósil. A diferencia de enfoques previos, que modelizan la extinción masiva como un único pulso instantáneo, el modelo BDSKY incorpora por primera vez la duración de estos eventos además de su intensidad, acorde con la definición paleontológica.
2. El modelo permite detectar períodos de tiempo (*time-slice*) en los que la tasa de extinción supera a la tasa de especiación, seguidos de períodos de recuperación, en los que la tasa de diversificación se hace positiva. La comparación de BDSKY con modelos bayesianos que modelizan la extinción masiva como un evento instantaneo ("pulso único") como CoMET demuestran que BDSKY proporciona mejores estimas de la tasa de extinción, pero que es sensible al número de especies en la filogenia y la intensidad o magnitud del evento de extinción. Las estimas más robustas se obtienen cuando la tasa de especiación se asume como constante.
3. El modelo BDSKY puede utilizarse para explorar la presencia de incrementos episódicos en la extinción "de fondo" durante la evolución de un linaje, cuando no existen hipótesis previas. Por ejemplo, en la detección de "cuellos de botella genéticos" en poblaciones de especies amenazadas, o para examinar el papel jugado por el cambio antropogénico (la "sexta extinción") en la biodiversidad actual.

4. Evidencia molecular de marcadores plastidiales y nuclear multicopia demuestra que el género de angiospermas africano *Camptoloma* Benth, un ejemplo paradigmático del patrón biogeográfico continental disyunto denominado "Rand Flora", se sitúa filogenéticamente como hermano de las tribus Buddlejeae y Teedieae, dentro de la familia Scrophulariaceae. *Camptoloma canariense*, endémica de Gran Canaria es hermana de *C. lyperiiflorum*, presente en Somalia y el sur de Arabia, mientras que *C. rotundifolium*, endémica de Namibia, en el sur de África, se reconstruye como su especie hermana.
5. El origen del género se sitúa en el suroeste de África en el Plioceno Inferior, hace 4.5 millones de años, mientras que la separación del grupo hermano (Teedieae-Buddlejeae) se remonta al Mioceno Inferior, hace 21 millones de años, cuando el clima era menos árido. La divergencia entre las tres especies implicó dispersión al norte y hacia el oeste, y eventos de vicarianza climática que fragmentaron la distribución ancestral como consecuencia de la progresiva aridificación del clima. *Camptoloma canariense* muestra una diversificación poblacional más reciente que las especies africanas, lo que sugiere una colonización reciente del archipiélago canario.
6. Los modelos de nicho climático proyectados sobre capas paleoclimáticas de los últimos 20 millones de años señalan la presencia de corredores climáticos durante el Mioceno Superior-Plioceno Inferior, entre el sur de África y el este de África (11-6 Ma), coincidente con un período de clima global más frío y húmedo. Previo al evento de calentamiento global del Mioceno Medio (3.5 Ma), se formaron también corredores entre el este y noroeste de África. Estos corredores desaparecen en las proyecciones actuales, lo que explica el grado de aislamiento geográfico entre las especies “extant”.
7. Se propone un novedoso enfoque en macroecología y evolución que combina modelos de nicho climático (ENMs) proyectados sobre capas paleoclimáticas, con técnicas de inferencia biogeográfica basados en filogenias actuales de especies. Las distribuciones ancestrales inferidas en el análisis biogeográfico se utilizan para seleccionar un valor de umbral que permita transformar probabilidades de adecuación climática en valores de presencia/ausencia en modelos de nicho climático. Al contrario que otros métodos, es posible seleccionar distintos valores de umbral para proyecciones pasadas, actuales o futuras. Las proyecciones ENM permiten, a su vez, informar a la inferencia biogeográfica sobre la presencia de corredores o barreras climáticas en intervalos de tiempo evolutivo para los que no existe ninguna evidencia cladogenética (ramas largas in eventos de diversificación ancestral). Este tipo de enfoque resulta particularmente útil en linajes con una edad de divergencia que se remonta a millones de años pero contienen pocas especies, y en los que la información geográfica asociada es escasa o sesgada por el muestreo, como en el género Rand Flora *Camptoloma*.

8. Dentro del campo de los modelos de simulación computacional, se propone el desarrollo matemático de un nuevo enfoque basado en simulaciones hacia el presente espacialmente explícitas en un paisaje

de celdas bidimensional proyectado sobre capas paleoclimáticas, que permite incorporar el papel de procesos evolutivos como especiación y extinción en la generación de filogenias. El nuevo enfoque expande modelos existentes basados en dispersión y dinámica de nicho climático para integrar la señal evolutiva en la formación de patrones de distribución de la biodiversidad.

9. El modelo ha sido escrito en lenguaje de programación PASCAL e incorpora tres capas o evidencias de información sobre un paisaje 2D, dividido en celdas: tasas de diversificación en linajes; un modelo de dispersión dependiente de distancia, que incorpora eventos aleatorios de dispersión a larga distancia; y adecuación de las condiciones del paisaje: abióticas (precipitación y temperatura), y bióticas (densidad de especies) para la supervivencia del linaje. Las relaciones dinámicas entre las tres capas pueden darse como interacciones "dentro" de una celda (extinción, adecuación biótica y abiótica, y éxito de colonización) o "entre" celdas (dispersión, adecuación abiótica, especiación). El modelo asume la conservación del nicho climático y una tasa constante de global especiación y extinción dentro de las celdas, pero no necesariamente entre celdas.
10. A diferencia de modelos previos que usan una tasa de especiación y consideran los eventos de especiación como sucesos evolutivos independientes, estos están condicionados o relacionados a través de un patrón de relaciones evolutivas y tiempos de divergencia, lo que permite comparar las filogenias generadas en la simulación con filogenias empíricas observadas. La tasa de extinción local depende del ajuste entre las condiciones climáticas existentes, y la tolerancia del linaje (variables abióticas), así como de la densidad de especies presentes en cada celda (variables bióticas). Este nuevo modelo macroecológico de simulación puede utilizarse para explorar hipótesis sobre la correlación entre cambios ambientales y conservación del nicho climático con la formación de patrones de distribución evolutivamente conservados y disyuntos, como es el caso del patrón biogeográfico Rand Flora.

11. Enfrentados a la grave crisis de biodiversidad y a la amenaza del calentamiento global como consecuencia de la actividad humana, es crucial desarrollar nuevas herramientas analíticas que permitan desentrañar la respuesta evolutiva de las especies en el pasado, como persistencia geográfica o migración de linajes, así como incrementar la precisión de nuestras predicciones para futuros cambios climáticos.

REFERENCE LIST

REFERENCE LIST

Adams C.G., Gentry A.W., Whybrow P.J. 1983. Dating the terminal Tethyan event. In Meulenkamp J. E. (Ed.), Reconstruction of marine paleoenvironments, Vol. 30. Utrecht Micropaleontological Bulletins Utrecht (pp. 273–298). Utrecht, the Netherlands: Utrecht University.

Al-Quaddi N., Vessella F., Stephan J., Al-Eisawi D., Schirone B. 2016. Current and future suitability areas of kermes oak (*Quercus coccifera* L.) in the Levant under climate change. Regional Environmental Change. 17:143-156.

Alfaro M.E., Santini F., Brock C., Alamillo H., Dornburg A., Rabosky D.L., Carnevale G., Harmon, L.J. 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. Proceedings of the National Academy of Science USA. 106:13410-13414.

Allouche O., Tsoar A., Kadmon R. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). Journal of Applied Ecology. 43:1223-1232.

Alroy J., Koch P.L., Zachos J.C. in Erwin D.H., Wing S.L. 2000. Eds. In Deep Time: Paleobiology's Perspective. The Paleontological Society, Lawrence, KS. 26:259-288.

Antonelli A., Sanmartín, I. 2011. Mass extinction, gradual cooling, or rapid radiation: Reconstructing the spatiotemporal evolution of the ancient angiosperm genus *Hedyosmum* (Chloranthaceae) using empirical and simulated approaches. Systematic Biology. 60:596-615.

Araújo M.A., Luoto M. 2007. The importance of biotic interactions for modelling species distributions under climate change. Global Ecology and Biogeography. 16:743-753.

Araújo M.B., New M. 2007. Ensemble forecasting of species distributions. Trends in Ecology and Evolution. 22:42-47.

Araújo M.B., Ferri-Yáñez F., Bozinovic F., Marquet P.A., Valladares F., Chown S.L. 2013. Heat freezes niche evolution Ecology Letters. 16:1206-1219.

Baillie J.E.M., Collen B., Amin R., Akcakaya H.R., Butchart S.H.M., Brummitt N., Meagher T.R., Ram M., Hilton-Taylor C., Mace G.M. 2008. Toward monitoring global biodiversity. Conservation Letters. 1:18-26.

Barnosky A.D., Matzke N., Tomiya S., Wogan G.O.U., Swartz B., Quental T.B., Marshall C., McGuire J.L., Lindsey E.L., Maguire K.C., Mersey B., Ferrer E.A. 2011a. Has the Earth's sixth mass extinction already arrived? Nature. 471:51-57.

Barnosky A.D., Carrasco M.A., Graham R.W. 2011b. Collateral mammal diversity loss associated with late Quaternary megafaunal extinctions and implications for the future. Geological Society, London, Special Publications. 358:179-189.

Basille M., Calenge C., Marboutin E., Andersen R., Gaillard, J.M. 2008. Assessing habitat selection using multivariate statistics: Some refinements of the ecological-niche factor analysis. Ecological Modelling. 211: 233-240.

Beaulieu J.M., O'Meara B.C. 2015. Extinction can be estimated from moderately sized molecular phylogenies. Evolution. 69:1036-1043.

Beerling D.J., Berner R.A., Mackenzie F.T., Harfoot M.B., Pyle J.A. 2009. Methane and the CH4 related greenhouse effect over the past 400 million years. American Journal of Science. 309:97-113.

Beerling D.J., Fox A., Stevenson D.S. Valdes P.J. 2011 Enhanced chemistry-climate feedbacks in past greenhouse worlds. Proceedings of the National Academy of Science USA. 108:9770-9775.

Beerling D.J., Taylor L.L., Bradshaw C.D.C., Lunt D.J., Valdes P.J., Banwart S.A., Pagani M., and Leake J.R. 2012. Ecosystem CO_2 starvation and terrestrial silicate weathering: mechanisms and

global-scale quantification during the late Miocene. Journal of Ecology. 100: 31-41.

Benkman C.W. 1995. The impact of tree squirrels (*Tamiasciurus*) on Limber Pine seed dispersal adaptations. Evolution. 49:585-592.

Benton M.J. 2009. The red queen and the court jester: species diversity and the role of biotic and abiotic factors through time. Science. 323:728-732.

Böhme M., Ilg A., Winklhofer 2008. Late Miocene "washhouse" climate in Europe. Earth and Planetary Science Letters. 275:393-401.

Boone R.B. 2010. Simulating species richness using agents with evolving niches, with an example of Galápagos plants. International Journal of Ecology. 2010:doi: 10.1155/2010/150606.

Bouckaert R., Heled J., Kühnert D., Vaughan T., Wu C-H., Xie D., Suchard M.A., Rambaut A., Drummond A.J. 2014. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. PLoS Computational Biology. 10:e1003537.

Brooks D.R. 1990. Parsimony analysis in historical biogeography and coevolution: methodological and theoretical update. Systematic Zoology. 39:14-30.

Brooks D.R., McLennan D.A. 1991. Phylogeny, ecology, and behavior: a research program in comparative biology. Chicago: University of Chicago Press.

Brown J.H., Lomolino M.V. 1998. Biogeography. Sinauer Associates, Sunderland, MA.

Bruch A.A., Utescher T., Mosbrugger V., Gabrielyan I., Ivanov D.A. 2006. Late Miocene climate in the circum-Alpine realm a quantitative analysis of terrestrial paleofloras. Palaegeography, Palaeoclimatology, Palaeoecology. 238:270-280.

Carboni M., Guéguen M., Barros C., Georges D., Boulangeat I., Douzet R., Dullinger S., Klonner G., van Kleunen M., Essl F., Bossdorf O., Hæuser E., Talluto M.V., Moser D., Block S., Conti L., Dullinger I., Münkemüller T., Thuiller W. 2018. Simulating plant invasion dynamics in mountain ecosystems under global change scenarios. Global Change Biology. 24: e289-e302.

Cerling T.E. 1984. The stable isotopic composition of modern soil carbonate and its relationship to climate. Earth and Planetary Science Letters. 71:229-240.

Cerling T.E., Quade J., Wang Y., Bowman J.R. 1989. Carbon isotopes in soils and paleosols as paleoecolgic indicators. Nature. 341:138-139.

Cerling T.E. 1992. Development of grasslands and savannas in East Africa during the Neogene. Palaeogeography, Palaeoclimatology, Palaeoecology (Global and Planetary Change Section). 97:241-247.

Cerling T.E., Harris J.M., MacFadden B.J., Leakey M.G., Quade J., Elsenmann V., Ehleringer J.R. 1997. Global vegetation change through the Miocene/Pliocene boundary. Nature. 389:153-158.

Chau T.H., Rahfeldt W.A., Olmstead R.G. 2018. Comparison of taxon-specific versus general locus sets for targeted sequence capture in plant phylogenomics. Applications in Plant Sciences. 6:e1032.

Chefaoui R.M., Lobo J.M. 2008. Assessing the effects of pseudo-absences on predictive distribution model performance. Ecological Modelling. 210:478-486.

Clark J.D., Hayes S.G., Pledger J.M. 1998. A female black bear denning habitat model using a geographic information system. Ursus. 10:181-185.

Cole M.M. 1986. The Savannas: Biogeography and geobotany. Academic Press, London: 438pp.

Condamine F.L., Rolland J., Morlon H. 2013. Macroevolutionary perspectives to environmental changes. Ecology Letters. 16:72-85.

Condamine F.L., Antonelli A., Lagomarsino L.P., Hoorn C., Liow L.H. 2018. Teasing apart mountain uplift, climate change and biotic drivers of species diversification. Mountains, Climate and Biodiversity. 18:257-272.

Connolly S.R. 2005. Process-based models of species distributions and the mod-domain effect. The American Naturalist. 166:1-11.

Copeland P. in Ruddiman W.F. 1997. Eds. Tectonic Uplift and Climate Change. Plenum, New York. 20-40.

Cracraft J. 1994. Species-diversity, biogeography, and the evolution of biotas. American Zoologist. 34:33-47.

Craven P., Craven D. 2000. The flora of the Brandberg, Namibia. Cimbebasia Memoir. 9:49-67.

Crisp M.D., Cook L.G. 2009. Explosive radiation of cryptic mass extinction? Interpreting signatures in molecular phylogenies. Evolution. 63:2257-2265.

De Maio N., Wu C-H., O'Reilly K.M., Wilson D. 2015. New routes to phylogeography: A Bayesian structured coalescent approximation. PLoS Genetics. 11:e1005421.

De Marco Jr P., Diniz-Filho J.A.F., Bini L.M. 2008. Spatial analysis improves species distribution modelling during range expansion. Biology Letters. 4:577-580.

De Vos J.M., Joppa L.N., Gittleman J.L., Stephens P.R., Pimm S.L. 2015. Estimating the normal background rate of species extinction. Conservation Biology. 29:452-462.

De Wit M.C.J., Bamford M.K. 1993. Fossil wood from the Brandvlei area, Bushmanland as an indication of palaeoenvironmental changes during the Cainozoic. Palaeontologia Africana 30:81-89.

Dechampes R. 1984. Evidence of bush fires during the Plio-Pleistocene in Africa (Omo and Sahabi) with the aid of fossil woods. Palaeoecology of Africa. 16:291-294.

Dechampes R., Senut B., Pickford M. 1992. Fruits fossiles pliocènes et pléistocènes du Rift occidental ougandais. Signification paléoenvironnementale. Comptes rendus de l'Académie des Sciences Paris, Series. II. 314:325-331.

Deines P. 1980. The isotopic composition of reduced organic carbon. In: P. Fritz and J.C. Fontes (Editors), Handbook of Environmental Geochemistry. 1. The Terrestrial Environment. Elsevier, Amsterdam. 329-406.

deMenocal P.B. 2004. African climate change and faunal evolution during the Pliocene-Pleistocene. Earth and Planetary Science Letters. 220:3-24.

Diniz-Filho J.A.F., Bini L.M. 2008. Macroecology, global change and the shadow of forgotten ancestors. Global Ecology and Biogeography. 17:11-17.

Diniz-Filho J.A.F., Ferro V.G., Santos T., Nabout J.C., Dobrovolski R., De Marco Jr. P. 2010. The three phases of the ensemble forecasting of niche models: geographic range and shifts in climatically suitable areas of *Utetheisa ornatrix* (Lepidoptera, Arctiidae). Revista Brasileira de Entomologia. 54:339-349.

Domina G., Marino P., Spadaro V., Raimondo F.M. 2012. Vascular flora evolution in the major Mediterranean islands. Biodiversity Journal. 3:337-342.

Donoghue M.J., Sanderson M.J. 2015. Confluence, synnovation, and depauperons in plant diversification. New Phytologist. 207:260-274.

Dormann C.F., McPherson J.M., Araújo M.B., Bivand R., Bolliger J., Carl G., Davies R.G., Hirzel A., Jetz W., Kissling W.D., Kühn I., Ohlemüller R., Peres-Neto P.R., Reineking B., Schröder B., Schurr F.M., Wilson R. 2007. Methods to account for spatial autocorrelation in the analysis of distributional species data: a review. Ecography. 30:609-628.

Dorn A., Musilová Z., Platzer M., Reichwald K., Cellerino A. 2014. The strange case of East African annual fishes: aridification correlates with diversification for a savannah aquatic group. BMC Evolutionary Biology. 14:210-242.

Eckert A.J., Hall B.D. 2006. Phylogeny, historical biogeography, and patterns of diversification for *Pinus* (Pinaceae): Phylogenetic tests of fossil-based hypotheses. Molecular Phylogenetics and Evolution. 40:166-182.

Ehleringer J.R. 1988. Carbon isotope rations and physiological processes in aridland plants. In: P.W. Rundel, J.R. Ehleringer and K.A. Nagy (Editors), Applications of Stable Isotopic Ratios to Ecological Research. Springer, New York, NY. 41-54.

Epstein S., Buchsbaum R., Lowenstam H.A., Urey H.C. 1953. Revised carbonate-water isotopic temperature scale. Geological Society of America Bulletin. 64:1315-1326.

Engler R., Guisan A., Rechsteiner L. 2004. An improved approach for predicting the distribution

of rare and endangered species from occurrence and pseudo-absence data. Journal of Applied Ecology. 41:263-274.

Evans M.E.K., Smith S.A., Flynn R.S., Donoghue M.J. 2009. Climate, niche evolution, and diversification of the "bird-cage" evening primroses (*Oenothera*, Sections *Anogra* and *Kleinia*). The American Naturalist. 173:225-240.

Ezard T.H.G., Aze T., Pearson P.N., Purvis A. 2011. Interplay Between Changing Climate and Species' Ecology Drives Macroevolutionary Dynamics. Science. 332:349-351.

Farjon A., Ortiz Garcia S. 2005. The early development of ovuliferous cones in Cupressaceae *s. lat* –a survey of the genera. Pp. 27-46 in A. Farjon, ed. A monograph of Cupressaceae and *Sciadopitys*. Royal Botanic Gardens, Kew, Richmond, Surrey, U.K.

Farrell B.D., Mitter C., Futuyma D.J. 1992. Diversification at the plant-insect interface. Bioscience. 42:34-42.

Fielding A.H., Bell J.F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation. 24:38-49.

Futuyma D.J. 1998. Evolutionary Biology, third edition Sinauer, Sunderland, MA.

García-Aloy S., Vitales D., Roquet C., Sanmartín I., Vargas P., Molero J., Kamau P., Aldasoro J.J., Alarcón M. 2017 North-west Africa as a source and refuge area of plant biodiversity: a case study on *Campanula kremeri* and *Campanula occidentalis*. Journal of Biogeography. 44:2057-2068.

Gotelli N.J., Anderson M.J., Arita H.T., Chao A., Colwell R.K., Connolly S.R., Currie D.J., Dunn R.R., Graves G.R., Green J.L., Grytnes J.-A., Jiang Y.-H., Jetz W., Lyons S.K., McCain C.M., Magurran A.E., Rahbek C., Rangel T.F.L.V.B., Soberón J., Webb C.O., Willig M.R. 2009. Patterns and causes of species richness: a general simulation model for macroecology. Ecology Letters. 12:873-886.

Gould S.J. 1994. Tempo and mode in the macroevolutionary reconstruction of Darwinism". PNAS. 91:6764-6771.Griffin D.L. 2002. Aridity and humidity: two aspects of the late Miocene climate of North Africa and the Mediterranean. Palaeogeography, Palaeoclimatology, Palaeoecology. 182:65-91.

Griffin D.L. 2002. Aridity and humidity: two aspects of the late Miocene climate of North Africa and the Mediterranean. Palaeogeography, Palaeoclimatology, Palaeoecology. 182:65-91.

Grytnes J.A. 2003 Ecological interpretation of the mid-domain effect. Ecology Letters 6:883-888.

Guisan A., Rahbek C. 2011. SESAM – a new framework integrating macroecology and species distribution models for predicting spatio-temporal patterns of species assemblages. Journal of Biogeography. 38:1433-1444.

Hannisdal B., Peters S.E. 2011. Phanerozoic Earth system evolution and marine biodiversity. Science. 334:1121-1124.

Hansen J., Sato M., Ruedy R., Lo K., Lea D.W., Medina-Elizade M. 2006. Global temperature change. Proceedings of the National Academy of Sciences. USA. 103:288-14 293.

Harvey P.H., May R.M., Nee S. 1994. Phylogenies Without Fossils. Evolution. 48:523-529.

Harzhauser M., Kroh A., Mandic O., Piller W.E., Göhlich U., Reuter M., and Berning, B. 2007. Biogeographic responses to geodynamics: A key study all around the Oligo-Miocene Tethyan Seaway. Zoologischer Anzeiger. 246:241-256.

Hassell M.P., Comins H.N., May R.M. 1991. Spatial structure and chaos in insect population dynamics. Letters to Nature. 353:255-158.

Hawkins B.A., Diniz-Filho J.A.F., Weis A.E. 2005 The mid-domain effect and diversity gradients: is there anything to learn. The American Naturalist 166:140-143.

Huang D., Goldberg E.E, Roy K. 2015. Fossils, phylogenies, and the challenge of preserving evolutionary history in the face of anthropogenic extinctions. Proceedings of the National Academy of Sciences. USA. 112:4909-4914.

Hengl T., Sierdsema H., Radović A., Dilo A. 2009. Spatial prediction of species' distributions from

occurrence-only records: combining point pattern analysis, ENFA and regression-kriging. Ecological Modelling. 220:3499-2511.

Hilliard O. 1994. The Manuleae: A Tribe of Scrophulariaceae. Edinburgh University Press, Edinburgh.

Hirzel A.H., Hausser J., Chessel D., Perrin N. 2002 Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? Ecology 83:2027-2036.

Hoffmann M., Hilton-Taylor C., Angulo A., Böhm M., Brooks T., Butchart S., Carpenter K., Chanson J., Collen B., Cox N., Darwall W., Dulvy N., Harrison L., Katariya V., Pollock C., Quader S., Dewhurst-Richman N., Rodrigues A., Tognelli M., Stuart S. 2010. The Impact of Conservation on the Status of the World's Vertebrates. Science. 330:1503-1509.

Höhna S., Stadler T., Ronquist F., Britton T. 2011. Inferring speciation and extinction rates under different sampling schemes. Molecular Biology and Evolution. 28:2577-2589.

Höhna S., May M.R., Moore B.R. 2015. TESS: An R package for efficiently simulating phylogenetic trees and performing Bayesian inference of lineage diversification rates. Bioinformatics. 32:789-791.

Höhna S., Landis M.J., Heath T.A., Boussau B., Lartillot N., Moore B.R., Huelsenbeck J.P., Ronquest F. 2016. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification Langauge. Systematic Biology. 65:726-736.

Hughes C.E., Eastwood R. 2006. Island ratification on a continental scale: exceptional rates of plant diversification after uplift of the Andes. Proceedings of the National Academy of Science USA. 103:10334-10339.

IPCC 2001. Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press.

Jablonski D. 2008. Extinction and spatial dynamics of biodiversity. Proceedings of the National Academy of Science USA. 105:11528-11535.

Jetz W., Rabnek C. 2001 Geometric constraints explain much of the species richness pattern in African birds. National Academy of Sciences 98:5661-5666.

Jiménez-Valverde A., Lobo J. M. 2007. Threshold criteria for conversion of probability of species presence to either-or presence-absence. Acta Oecologica. 31:361-369.

Judson O.P. 1994 The rise of the individual-based model in ecology. Proceedings of the National Academy of Science USA. 98:5661-5666.

Kapli P., Lymberakis P., Crochet P.A., Geniez P., Brito J.C., Almutairi M., Schmitz A., Wilms T., Pouyani N.R., Poulakakis N. 2015. Historical biogeography of the lacertid lizard Mesalina in North Africa and the Middle East. Journal of Biogeography. 42:267-279.

Kornhall P., Heidari N., Bremer B. 2001. Selagineae and Manuleeae, two tribes or one? Phylogenetic studies in the Scrophulariaceae. Plant Systematics and Evolution. 228:199-218.

Kozak, K.H., Graham C.H., Wiens J.J. 2008. Integrating GIS-based environmental data into evolutionary biology. Trends in Ecology & Evolution. 23:141-148.

Landis J.L., Matzke N.J., Moore B.R., Huelsenbeck J.P. 2013. Bayesian analysis of biogeography when the number of areas is large. Systematic Biology. 62:789-804.

Landis J.B., Soltis D.E., Li Z., Marx H.E., Barker M.S., Tank D.C., Soltis P.S. 2018. Impact of whole-genome duplication events on diversification rates in angiosperms. American Journal of Botany. 105:348-363.

Latham E.L., Ricklefs R.E. 1993. Continental comparisons of temperate-zone tree species diversity. Species diversity in ecological communities: historical and geographical perspectives. University of Chicago Press, Chicago. (edition. by Ricklefs R.E. and Schluter D.):294-314.

Laurent S., Robinson-Rechavi M., Salamin N. 2015. Detecting patterns of species diversification in the presence of both rate shifts and mass extinctions. BMC Evolutionary Biology. 15:doi: 10.1186/s12862-015-0432-z.

Lawver L.A., Gahagan L.M. in Crowley T.J., Burke K.G. 1998. Eds. Tectonic Boundary Conditions for Climate Reconstructions. Oxford University Press, New York. 39:212-226.

Leakey M.G., Feibel C.S., Bernor R.L., Harris J.M., Cerling T.E., Stewart K.M., Storrs G.W., Walker A., Werdelin L., Winkler, A.J. 1996. Lothagam: A record of faunal change in the late Miocene of east Africa. Journal of Vertebrate Paleontology. 16:556-570.

Lear C.H., Rosenthal Y., Wright J.D. 2003. The closing of a seaway: ocean water masses and global climate change. Earth and Planetary Science Letters. 210:425-436.

Lemey P., Rambaut A., Drummond A.J., Suchard M.A. 2009. Bayesian phylogeography finds its roots. PLoS Computational Biology. 5:e1000520. doi:10.1371.

Leslie A.B., Beautlieu J.M., Rai H.S., Crane P.R., Donoghue M.J., Mathews, S. 2012. Hemisphere-scale differences in conifer evolutionary dynamics. Proceedings of the National Academy of Science USA. 109:16217-162.

Levins R. 1969. Some demographic and genetic consequences of environmental heterogeneity for biological control. Bulletin of the Entomological Society of America. 15:237-240.

Liu C., Berry P.M., Dawson T.P., Pearson R.G. 2005. Selecting thresholds of occurrence in the predictions of species distributions. Ecography. 28:385-393.

Liu H., Li S., Ugolini A., Momtazi F., Hou Z. 2018. Tethyan closure drove tropical marine biodiversity: Vicariant diversification of intertidal crustaceans. Journal of Biogeography. 45:941-951.

Lobo J.M., Jiménez-Valverde A., Real R. 2008. AUC: a misleading measure of the performance of predictive distribution models. Global Ecology and Biogeography. 17:145-151.

Lourens L.J., Sluijs A., Kroon D., Zachos J.C., Thomas E., Röhl U., Bowles J., Raffi I. 2005. Astronomical pacing of late Palaeocene to early Eocene global warming events. Nature. 435:1083-1087.

Lunt D.J., Flecker R., Valdes P.J., Salzmann U., Gladstone R., Haywood A.M. 2008. A methodology for targeting palaeo proxy data acquisition: a case study for the terrestrial late Miocene. Earth and Planetary Science Letters. 271:53–62.

Mabberley D.J. 2008. Mabberley's plant book: a portable dictionary of plants, their classifications, and uses. Cambridge University Press, New York.

Mace G.M., Collar N., Gaston K., Hilton-Taylor C., Akcakaya H.R., Leader-Williams N., Milner-Gulland E., Stuart S. 2008. Quantification of Extinction Risk: IUCN's System for Classifying Threatened Species. Conservation biology: The Journal of the Society for Conservation Biology. 22:1424-1442 (This paper explains the methodology used by the IUCN to assess the extinction risks of extant species).

Magallón S., Gómez-Acevedo S., Sánchez-Reyes L.L., Hernández-Hernández T. 2015. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. New Phytologist. 207:437-453.

Mairal M., Pokorny L., Aldasoro J.J., Alarcón M., Sanmartín I. 2015. Ancient vicariance and climate-driven extinction explain continental-wide disjunctions in Africa: the case of the Rand Flora genus *Canarina* (Campanulaceae). Molecular Evolution. 24:1335-1354.

Mairal M., Sanmartín I., Pellissier L. 2017. Lineage-specific climatic niche drives the tempo of vicariance in the Rand Flora. Journal of Biogeography. 44:doi: 10.1111/jbi.12930.

Mairal M., Caujapé-Castells J., Pellissier L., Jaén-Molina R., Álvarez N., Heuertz M., Sanmartín I. 2018. A tale of two forests: ongoing aridification drives population decline and genetic diversity loss at continental scale in Afro-Macaronesian evergreen-forest archipelago endemics. Annuals of Botany 122:1005-1017.

Mao K., Hao G., Liu J., Adams R.P., Milne R.I. 2010. Diversification and biogeography of *Juniperus* (Cupressaceae): variable diversification rates and multiple intercontinental dispersals. New Phytologist. 188:254-272.

Martínez-Meyer, E., Peterson A.T., Hargrove W.W. 2004. Ecological niches as stable distributional

constraints on mammal species, with implications for Pleistocene extinctions and climate change projections for biodiversity. Global Ecology and Biogeography. 13:305-314.

Martínez-Meyer E., Peterson A.T. 2006. Conservatism of ecological niche characteristics in North American plant species over the Pleistocene-to-recent transition. Journal of Biogeography. 33:1779-1789.

May M.R., Höhna S., Moore B.R. 2016. A Bayesian approach for detecting the impact of mass-extinction events on molecular phylogenies when rates of lineage diversification may vary. Methods in Ecology and Evolution. 7: 947-959.

McDonald K.A., Brown J.H. 1992. Using montane mammals to model extinctions due to global change. Conservation Biology. 6:409-415.

Meseguer A.S., Lobo J.M., Ree R., Beerline D.J., Sanmartín I. 2015. Integrating fossils, phylogenies, and niche models into biogeography to reveal ancient evolutionary history: The case of *Hypericum* (Hypericaceae). Systematic Biology. 64:215-232.

Meseguer A.S., Lobo J.M., Cornuault J., Beerling D., Ruhfel B.R., Davis C.C., Jousselin E., Sanmartín I. 2018. Reconstructing deep-time palaeoclimate legacies in the clusioid Malpighiales unveils their role in the evolution and extinction of the boreotropical flora. Global Ecology and Biogeography. 27:616-628.

Metcalf J.L., Prost S., Nogués-Bravo D., DeChaine E.G., Anderson C., Batra P., Araújo M.B., Cooper A., Guralnick R.P. 2014. Integrating multiple lines of evidence into historical biogeography hypothesis testing: A *Bison bison* case study. Proceedings of the Royal Society B. 281:20132782.

Miller K.G., Kominz M.A., Browning J.V., Wright J.D., Mountain G.S., Katz M.E., Sugarman P.J., Cramer B.S., Christie-Blick N., Peker S.F. 2005. The Phanerozoic record of global sea-level change. Science. 312:1293-1298.

Moen D., Morlon H. 2014. Why does diversification slow down? Trends in Ecology & Evolution. 29:190-197.

Monserud R.A., Leemans R. 1992. Comparing global vegetation maps with the Kappa statistic. Ecological Modelling. 62:275-293.

Morlon H., Parsons T.L., Plotkin J.B. 2011. Reconciling molecular phylogenies with the fossil record. Proceedings of the National Academy of Science USA. 108:16327-16332.

Mosbrugger V., Utescher T., Dilcher D.L. 2005. Cenozoic continental climate evolution of Central Europe. Proceedings of the National Academy of Sciences. USA. 102:14964-14969.

Nee S.N., May R.M., Harvey P.H. 1994. The reconstructed evolutionary process. Philosophical Transactions of the Royal Society B. 344:305-311.

Otto-Bliesner B.L., Marshall S.J., Overpeck J.T., Miller G.H., Hu A., CAPE Last Interglacial Project members 2006. Simulating arctic climate warmth and icefield retreat in the last interglaciation. Science 311:1751-1753.

Owens J.N., Catalano G.L., Aitken-Christie J. 1997. The reproductive biology of Kauri (*Agathis australis*). IV. Late embryology, histochemistry, cone and seed development. International Journal of Plant Sciences. 158:395-407.

Owens J.N., Takaso T., Runions C.J. 1998. Pollination in conifers. Trends in Plant Science. 3:479-485.

Owens H.L., Campbell L.P., Dornak L.L. Saupe E.E., Barve N., Soberón J., Ingenloff K., Lira-Noriega A., Hensz C.M., Myers C.E., Peterson A.T. 2011. Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas. Ecological Modelling. 263:10-18.

Oxelman B., Kornhall P., Olmstead R.G., Bremer B. 2005. Further disintegration of Scrophulariaceae. TAXON. 54:411-425.

Pearson R.G., Thuiller W., Araújo M.B., Maerinez-Meyer E., Brotons L., McClean C., Miles L., Segurado P., Dawson T.P., Lees D.C. 2006. Moldel-based uncertainty in species range prediction. Journal of Biogeography. 33:1704-1711.

Pearson R.G., Raxworthy C.J., Nakamura M., Peterson A.T. 2007. Predicting species distributions from small numbers of occurrence

records: A test case using cryptic geckos in Madagascar. Journal of Biogeography. 34:102-117.

Pereira H.M., Leadley P.W., Proença V., Alkemade R., Scharlemann J.P.W., Fernandez-Manjarrés J.F., Araújo M.B., Balvanera P., Biggs R., Cheung W.W.L., Chini L., Cooper H.D., Gilman E.L., Guénette S., Hurtt G.C., Huntington H.P., Mace G.M., Oberdorff T., Revenga C., Rodrigues P., Scholes R.J., Sumaila U.R., Walpole M. 2010. Scenarios for global biodiversity in the 21st century. Science. 330:1496-1501.

Peterson A.T. 2006. Use and requirements of ecological niche models and related distributional models. Biodiversity Informatics. 3:59-72.

Peterson A.T., Soberón J., Pearson R.G., Anderson R.P., Nakamura M., Martinez-Meyer E., Nakamura M., Miguel A. 2011. Ecological Niches and Geographical Distributions. Princeton University Press, New Jersey.

Pickford M., Senut B. 1999. Geology and palaeobiology of the Namib Desert, southwestern Africa. Volume 1: Geology and history of study. Memoirs of the Geological Survey of Namibia. 18:1-155.

Pickford M., Senut B. (Eds.) 2003. Geology and palaeobiology of the central and southern Namib. Volume 2: Palaeontology of the orange river valley, Namibia. Memoirs of the Geological Survey of Namibia. 19 398p.

Pickford M. 2008. Diversification of grazing mammals in southern and equatorial Africa during the Neogene and Quaternary. Memoir of the Geological Survey of Namibia. 529-537.

Pimm S., Raven P., Peterson A., Sekercioglu Ç.H., Ehrlich P.R. 2006. Human impacts on the rates of recent, present, and future bird extinctions. Proceedings of the National Academy of Sciences USA. 103:10941-10946.

Pimm S.L., Joppa L.N. 2015. How many plant species are there, where are they, and at what rate are they going extinct? Annuals of the Missouri Botanical Garden. 100:170-176.

Pirie M., Kandziora M.D., Nürk N.M., Le Maitre N.C., Mugrabide Kuppler A.L., Gehrke B., Oliver E.G.H., Bellstedt D.U. 2018. Leaps and bounds: geographical and ecological distance constrained the colonisation of the Afrotemperate by *Erica*. BioRxiv. 290791:doi: 10.1101/290791.

Plana V. 2004. Mechanisms and tempo of evolution in the African Guineo-Congolian rainforest. Philosophical Transactions of the Royal Society B. 359:1585-1594.

Plotnik, R.E., Kenig F., Scott A.C., Glasspool I.J., Eble C.F., Lang W.J. 2009. Pennsylvanian paleokarst and cave fills from Northern Illinois, USA: a window into Late Carboniferous environments and landscapes. Palaios. 24:627-637.

Pokorny L., Riina R., Mairal M., Meseguer A.S., Culshaw V., Cendoya J., Serrano M., Carbajal, R., Ortiz S., Heuertz M., Sanmartín I. 2015. Living of the edge: timing of Rand Flora disjunctions congruent with ongoing aridification in Africa. Frontiers in Genetics. 6:doi: 10.3389/fgene.2015.00154.

Pratt D.J., Greenway P.J., Gwynne M.D. 1966. A classification of East African rangeland with an appendix on terminology. Journal of Applied Ecology. 3:369-382.

Purvis A. 2008. Phylogenetic approach to the study of extinction. Annual Review of Ecology, Evolution, and Systematics. 39:301-319.

Pyron R.A., Burbrink F.T. 2012. Trait-dependent diversification and the impact of paleontological data on evolutionary hypothesis testing in New World rattlesnakes (tribe Lampropeltini). Journal of Evolutionary Biology. 25:497-508.

Quade J., Cerling T.E., Bowman J.R. 1989a. Development of Asian monsoon revealed by marked ecological shift during the latest Miocene in northern Pakistan. Nature. 342:163-166.

Quade J., Cerling T.E., Bowman J.R. 1989b. Systematic variations in the stable carbon and oxygen isotopic composition of pedogenic carbonate along elevation transects in the southern Great Basin, USA. Geological Society of American Bulletin. 11:464 475.

Quental T.B., Marshall C.R. 2010. Diversity dynamics: Molecular phylogenies need the fossil record. Trends in Ecology and Evolution. 25:434-441.

Rabosky D.L. 2006. Likelihood methods for detecting temporal shifts in diversification rates. Evolution. 60:1152-1164.

Rabosky D.L., Lovette I.J. 2008. Explosive evolutionary radiations: Decreasing speciation or increasing extinction through time? Evolution. 62:1866-1875.

Rabosky D.L. 2010. Extinction rates should not be estimated from molecular phylogenies. Evolution. 64:1816-1824.

Rabosky D.L. 2014. Automatic Detection of Key Innovations, Rate Shifts, and Diversity-Dependence on Phylogenetic Trees. PLOS ONE. 9:e89543.

Rahbek C., Gotelli N.J., Colwell R.K., Entsminger G.L., Rangel T.F.L.V.B., Graves G.R. 2007. Predicting continental-scale patterns of bird species richness with spatially explicit models. Proceedings of the Royal Society B: Biological Sciences. 274:165-174.

Rangel T.F.L.V.B., Diniz-Filho J.A.F., Colwell R.K. 2007 Richness and evolutionary niche dynamics: a spatial pattern-oriented simulation experiment. American Society of Naturalists. 170:602-616.

Rannala B. 2002. Identifiably of parameters in MCMC Bayesian inference of phylogeny. Systematic Biology. 51:754-760.

Raup D.M. 1979. Size of the Permo-Triassic bottleneck and its evolutionary implications. Science. 206:217-218.

Ree R.H., Moore B.R., Webb C.O., Donoghue M.J. 2005. A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. Evolution. 59:2299-2311.

Ree R.H., Smith S.A. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. Systematic Biology. 57:4-14.

Richards C.L., Carstens B.C., Knowles L.L. 2007. Distribution modelling and statistical phylogeography: an integrative framework for generating and testing alternative biogeographical hypotheses. Journal of Biogeography. 34:1833-1845.

Ricklefs R.E., Schluter, D. 1993. Species diversity: regional and historical influences. In species diversity in ecological communities. University of Chicago Press. (edition Ricklefs R.E. and Schluter D.):350-363.

Romdal T.S., Araújo M.B., Rahbek C. 2013. Life on a tropical planet: niche conservatism and the global diversity gradient. Global Ecology Biogeography. 22:344-350.

Ronquist F., Sanmartín, I. 2011. Phylogenetic methods in biogeography. Annual Review of Ecology, Evolution, and Systematics. 42:441-464.

Royer D.L. 2006. CO_2-forced climate thresholds during the Phanerozoic. Geochimica et Cosmochimica Acta. 70:5665-5675.

Sanmartín I., Engho H., Ronquist F. 2001. Patterns of animal dispersal, vicariance and diversification in the Holarctic. Biological Journal of the Linnean Society. 73:345-390.

Sanmartín I., van Der Mark P., Ronquist F. 2008. Inferring dispersal: A Bayesian approach to phylogeny-based island biogeography, with special reference to the Canary Islands Journal of Biogeography. 35:428-449.

Sanmartín I., Anderson C.L., Alarcon M., Ronquist F., Aldasoro J.J. 2010. Bayesian island biogeography in a continental setting: The Rand Flora case. Biology Letters. 6:703-707.

Sanmartín I., Meseguer A.S. 2016. Extinction in phylogenetics and biogeography: from timetrees to patterns of biotic assemblage. Frontiers in Genetics. 7:35.

Sarmiento G. 1984. *The Ecology, of Neotropical Savannas,* Harvard University Press, Cambridge, MA. 235 pp.

Saupe E.E., Barve N., Owens H.L., Cooper J.C., Hosner P.A., Peterson A.T. 2017. Reconstructing ecological niche evolution when niches are incompletely characterized. Systematic Biology. 67:428-438.

Schefuss E., Schouten S., Jansen J.H.F., Sinnighe Damsté J.S., 2003. African vegetation controlled by tropical sea surface temperatures in the mid-Pleistocene period. Nature. 422:418-421.

Scott A.C., Chaloner W.G. 1983. The earliest fossil conifer from the Westphalian B of Yorkshire.

Proceedings of the Royal Society of London B. 220:163-182.

Ségalen L., Renard M., Lee-Thorp J.A., Emmanuel L., Le Callonnec L., de Rafélis M., Senut B., Pickford M., Melice J.-L. 2006. Neogene climate change and emergence of C4 grasses in the Namib, southwestern Africa, as reflected in ratite 13C and 18O. Earth and Planetary Science Letters. 244:725-734.

Ségalen L., de Rafélis M., Lee-Thorp J.A., Maurer A.-F., Renard M. 2008. Cathodoluminescence tools provide clues to depositional history in Miocene and Pliocene mammalian teeth. Palaeogeography, Palaeoclimatology, Palaeoecology. 266 :246-253.

Senut B., Pickford M., Wessels D. 1997. Panafrican distribution of Lower Miocene Hominoidea. Comptes rendus de l'Académie des sciences. 325:741–746.

Senut B., Pickford M., Ségalen L. 2009. Neogene desertification of Africa. Comptes Rendus Geoscience. 341:591-602.

Sepkoski Jr J.J. 1982. Mass extinctions in the Phanerozoic oceans: A review. Geological Implications of Impacts of Large Asteroids and Comets on the Earth, Leon T. Silver, Peter H. Schultz.

Sepulchre P., Ramstein G., Fluteau F., Schuster M., Tiercelin J.J., Brunet M. 2006. Tectonic uplift and Eastern Africa aridification. Science. 313:1419-1423.

Siepielski A.M., Benkman C.W. 2007a. Convergent patterns in the selection mosaic for two North American bird-dispersed pines. Ecological Monographs. 77:203-220.

Siepielski A.M., Benkman C.W. 2007b. Selection by a predispersal seed predator con- strains the evolution of avian seed dispersal in pines. Functional Ecology 21:611-618.

Siepielski A.M., Benkman C.W. 2008. A seed predator drives the evolution of a seed dispersal mutualism. Proceedings of the Royal Society of London B. 275:1917-1925.

Smith S.A., Donoghue M.J. 2010. Combining historical biogeography with niche modeling in the *Caprifolium* clade of *Lonicera* (Caprifoliaceae, Dipsacales). Systematic Biology. 59:322-341.

Stadler T. 2011a. Inferring speciation and extinction processes from extant species data. Proceedings of the National Academy of Science USA. 108:16145-16146.

Stadler T. 2011b. Mammalian phylogeny reveals recent diversification rate shifts. Proceedings of the National Academy of Science USA. 108:6187-6192.

Stadler T. 2011c. Simulating trees with a fixed number of extant species. Systematic Biology. 60:676-684.

Stadler T., Bokma F. 2012. Estimating Speciation and Extinction Rates for Phylogenies of Higher Taxa. Systematic Biology. 62:220-230.

Stadler T., Kühnert D., Bonhoeffer S., Drummond A.J. 2013. Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). Proceedings of the National Academy of Science USA. 110:228-233.

Storch D., Davis R.G., Zajíček S., Orme C.D.L., Olson V., Thomas G.H., Ding T.-S., Rasmussen R.C., Ridgely R.S., Bennett, P.M., Blackburn T.M., Owens I.P.F., Gaston K.J. 2006. Energy, range dynamics and global species richness patterns: Reconciling mid-domain effects and environmental determinants of avian diversity. Ecology Letters. 9:1308-1320.

Svenning J.-C., Eiserhardt W.L., Normand S., Ordonez A., Sandel B. 2015. The influence of palaeoclimate on present-day patterns in biodiversity and ecosystems. Annual Review of Ecology, Evolution, and Systematics. 46:551-572.

Tagliacollo V.A., Roxo F.F., Duke-Sylvester S.M., Oliveira C., Albert J.S. 2015. Biogeographical signature of river capture events in Amazonian lowlands. Journal of Biogeography. 42:2349-2362.

Takaso T., Owens J.N. 1995. Ovulate cone morphology and pollination in *Pseudotsuga* and *Cedrus*. International Journal of Plant Sciences. 156:630-635.

Thiede J., Winkler A., Wolf-Welling T., Eldholm O., Myhre A.M., Baumann K.H., Henrichs R., Stein R. 1998. Late Cenozoic history of the polar North Atlantic: results from ocean drilling. Quaternary Science Reviews. 17:185-208.

Tiffney B.H. 1985a. Perspectives on the origin of the floristic similarity between eastern Asia and eastern North America. Journal of the Arnold Arboretum. 66:73-94.

Tiffney B.H. 1985b. The Eocene North Atlantic land bridge: its importance in Tertiary and modern phytogeography of the Northern Hemisphere. Journal of the Arnold Arboretum. 66:243-273.

Thiv M., Thulin M., Hjertson M., Kropf M., Linder H.P. 2010. Evidence for a vicariant origin of Macaronesian-Eritreo/Arabian disjunctions in *Campylanthus* Roth (Plantaginaceae). Molecular Phylogenetics and Evolution. 54:607-616.

Thuiller W., Lavorel S., Sykes M.T., Araújo M.B. 2006. Using niche-based modelling to assess the impact of climate change on tree functional diversity in Europe. Diversity and Distributions. 12:49-60.

Thuiller W., Albert C., Araújo M.B., Berry P.M., Cabeza M., Guisan A., Hicker T., Midgley G.F., Paterson J., Schurr F.M., Sykes M.T., Zimmermann N.E. 2008. Prediction global change impacts on plant species' distributions: Future challenges. Perspectives in Plant Ecology, Evolution and Systematics. 9:137-152.

Thuiller W., Lafourcade B., Engler R., Araújo M.B. 2009. BIOMOD – a platform for ensemble forecasting of species distributions. Ecography 32:369-373.

Tomlinson P.B., Braggins J.E., Ratenbury J.A. 1991. Pollination drop in relation to cone morphology in Podocarpaceae: a novel reproductive mechanism. American Journal of Botany. 78:1289-1303.

Tomlinson, P. B., and T. Takaso. 2002. Seed cone structure in conifers in relation to development and pollination: a biological approach. Canadian Journal of Botany. 80:1250-1273.

Valente L.M., Savalainen V., Vargas P. 2010. Unparalleled rates of species diversification in Europe. Proceedings of the Royal Society B: Biological Sciences. 277:1489-1496.

Vaughan T., Drummond P., Drummond, A. 2012. Within-host demographic fluctuations and correlations in early retroviral infection. Journal of Theoretical Biology. 295:86-99.

Villaverde T., Pokorny L., Olsson S., Rincón-Barrado M., Johnson M.G., Gardner E.M., Wickett N.J., Molero J., Riina R., Sanmartín I. 2018. Bridging the micro- and macroevolutionary levels in phylogenomics: Hyb-Seq solves relationships from populations to species and above. New Phytologist. 220:636-650.

Waldron A. 2010. Lineages that cheat death: surviving the squeeze on range size. Evolution. 64: 2278-2292.

Wang X.-Q., Ran J.-H. 2014. Evolution and biogeography of gymnosperms. Molecular Phylogenetics and Evolution. 75:24-40.

Webb C.O., Ackerly D.D., McPeek M.A., Donoghue M.J. 2002. Phylogenies and community ecology. Annual Review of Ecology and Systematics. 33:475-505.

Wiens J.J. 2004. Speciation and ecology revisited: phylogenetic niche conservatism and the origin of species. Evolution. 58:193-197.

Wiens J.J., Donoghue M.J. 2004. Historical biogeography, ecology and species richness. TRENDS in Ecology and Evolution. 19:639-644.

Wiens J.J., Ackerly D.D., Allen A.P., Anacker B.L., Buckley L.B., Cornwell H.V., Damschen E.I., Davies T.J., Grytnes J.-A., Harrison S.P., Hawkins B.A., Holt R.D., McClain C.M., Stephens P.R. 2010. Niche conservatism as an emerging principle in ecology and conservation biology. Ecology Letters. 13:1310-1324.

Wiley E.O. 1981. Phylogenetics: The Theory and Practice of Phylogenetic Systematics. Wiley-Interscience.

Willig M.R., Kaufman D.M., Stevens R.D. 2003. Latitudinal gradients of biodiversity: pattern, process, scale, and synthesis. Annual Review of Ecology, Evolution, and Systematics. 34:273-309.

Willis K.J., MacDonald G.M. 2011. Long-term ecological records and their relevance to climate change predictions for a warmer world. Annual Review of Ecology, Evolution, and Systematics. 42:267-287.

Wolfe J.A. 1975. Some aspects of plant geography of the northern hemisphere during the Late cretaceous and Tertiary. Annals of the Missouri Botanical Garden. 62:264-279.

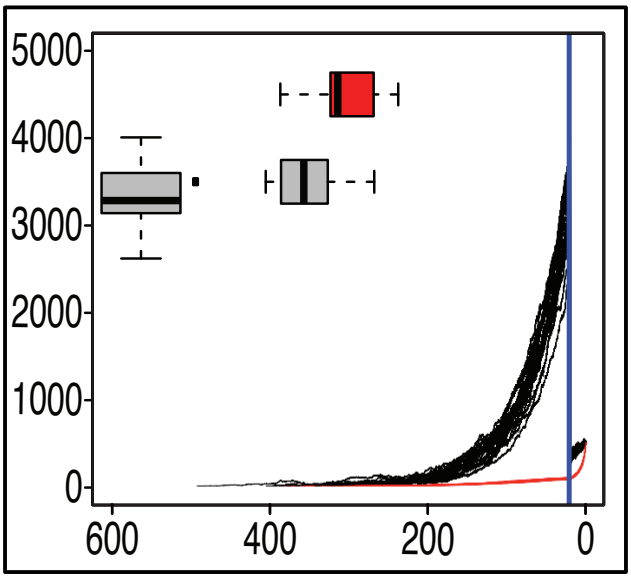
Zachos J., Pagani M., Sloan L., Thomas E., Bullups K. 2001. Trends, rhythms, and aberrations in global climate 65 Ma to Present. SCIENCE. 292:686-693.

Zachos J.C., Gerald R.D., Zeebe R.E. 2008. An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. NATURE. 451:279-283.

APPENDICES

CHAPTER 1: APPENDIX

Exploring the power of Bayesian birth-death skyline models to detect mass extinction events from phylogenies with only extant taxa



Supplementary Figures

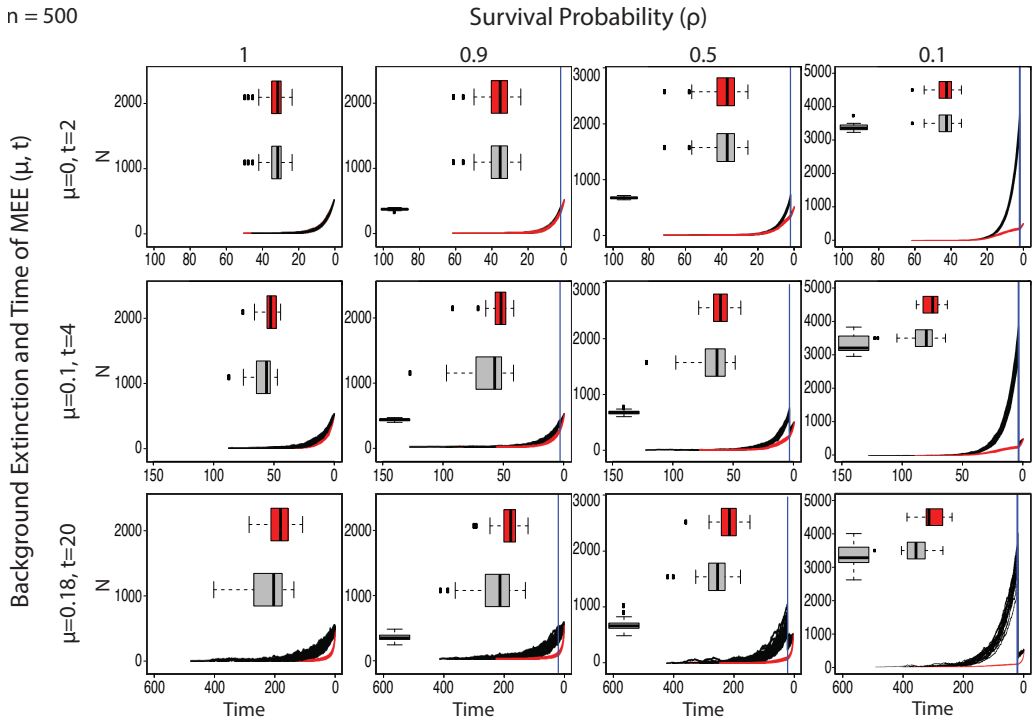


Figure S1: LTT plots of simulated phylogenies under different scenarios with varying values of μ and ρ. Black lines represent the LTT plot of the full (extinct and extant taxa) tree; the red lines show the reconstructed, extant-only tree LTT plot. Horizontal boxplots represent the variation in the age of the root for the full (grey) and reconstructed (red) trees. Vertical boxplots represent the variation in the number of pre-MEE lineages in the models with low mass extinction survival probability, ρ.

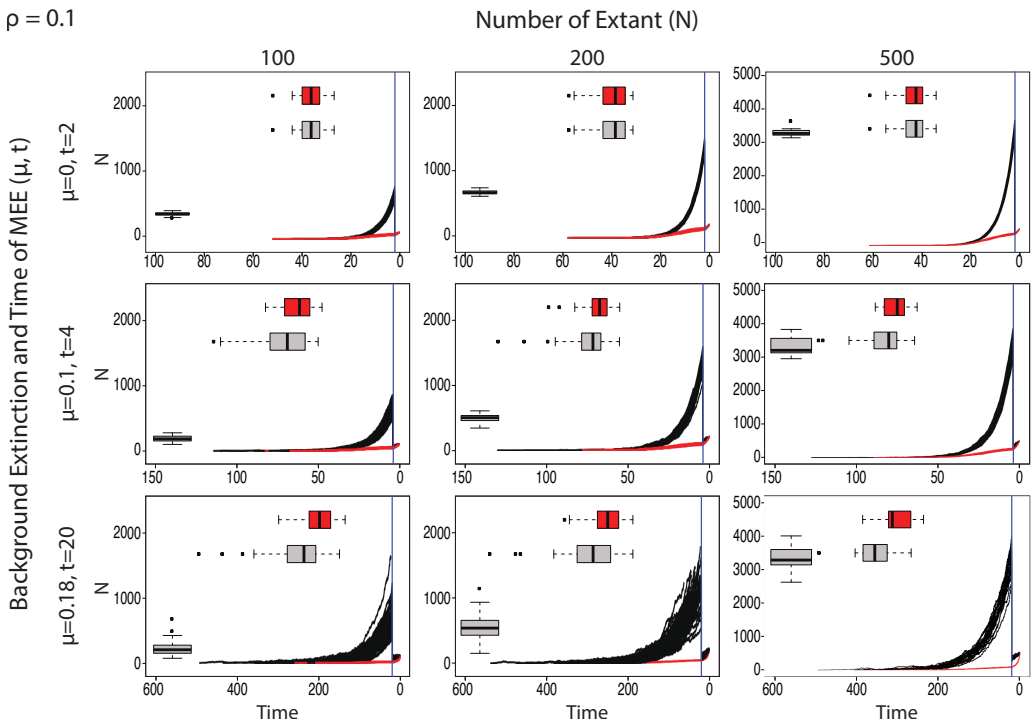


Figure S2: LTT plots of simulated phylogenies under different values of μ and number of extant taxa in reconstructed phylogeny (N). Black lines represent the LTT plot of the full (extinct and extant taxa) tree; the red lines show the reconstructed, extant-only tree LTT plot. All other conventions as in Figure S1.

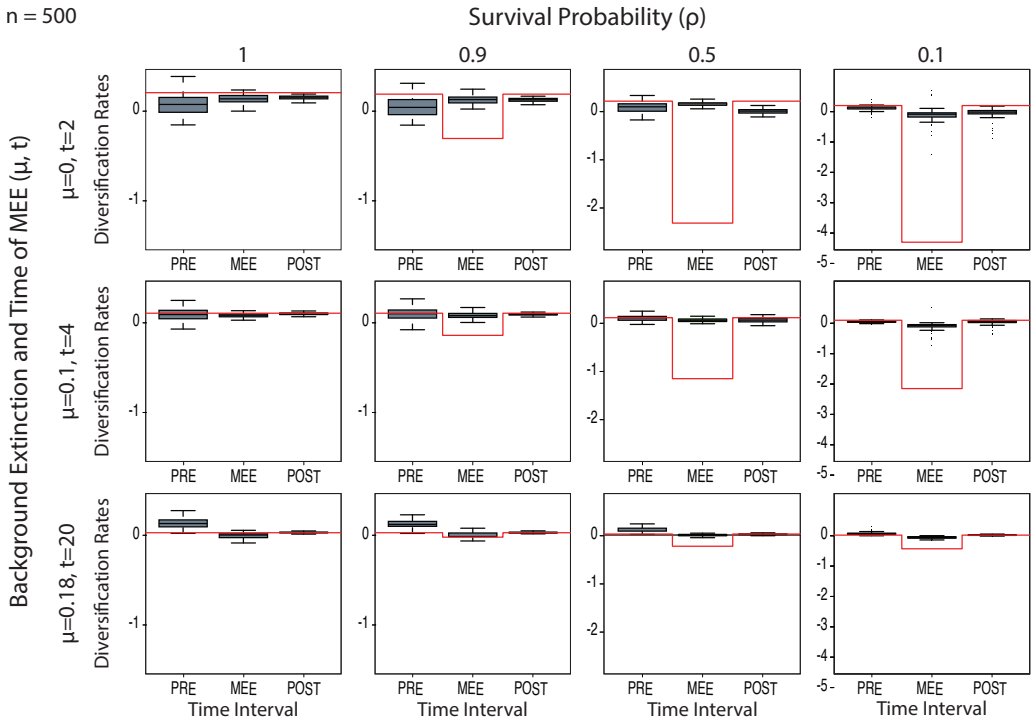


Figure S3: Detection of MEEs through interoperating changes in the magnitude of the diversification rate in Model A under varying levels of μ and mass extinction survival probability, ρ. All other conventions follow Figure 2.

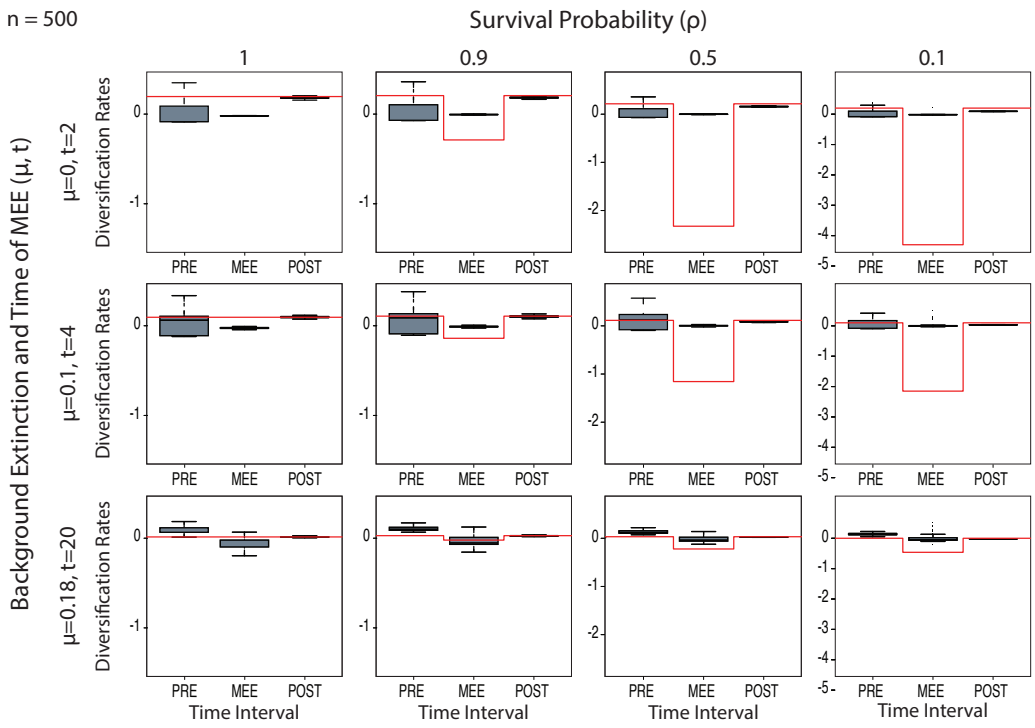


Figure S4: Detection of MEEs through interoperating changes in the magnitude of the diversification rate in Model B under varying levels of μ and mass extinction survival probability, ρ. All other conventions follow Figure 2.

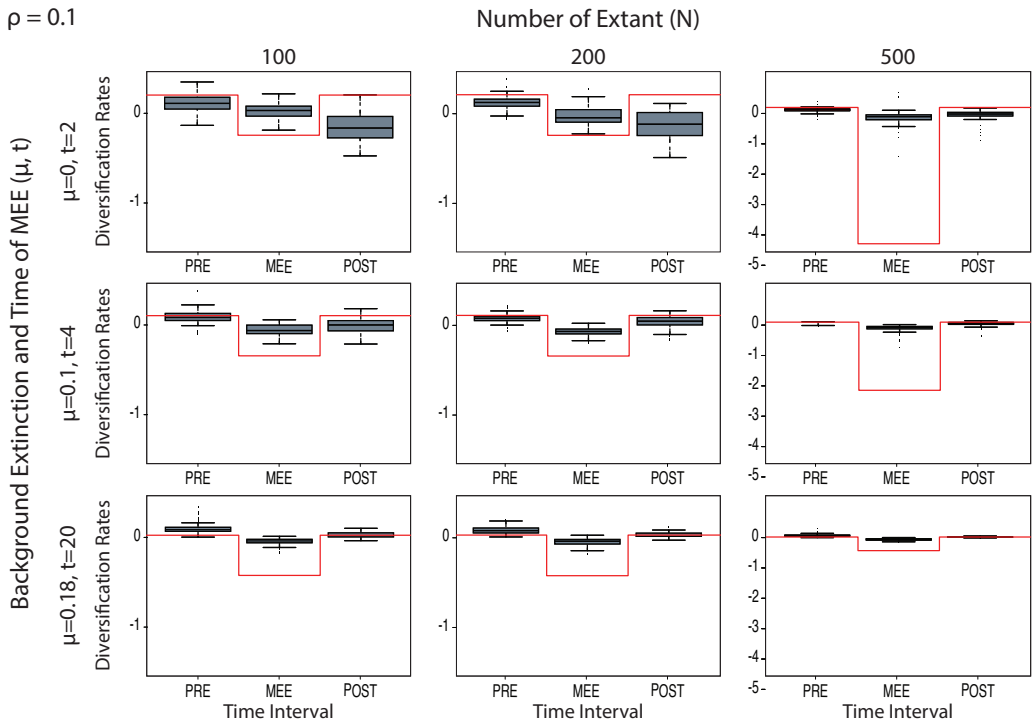


Figure S5: Detection of MEEs through interoperating changes in the magnitude of the diversification rate in Model A under varying levels of N (number of taxa). All other conventions follow Figure 2.

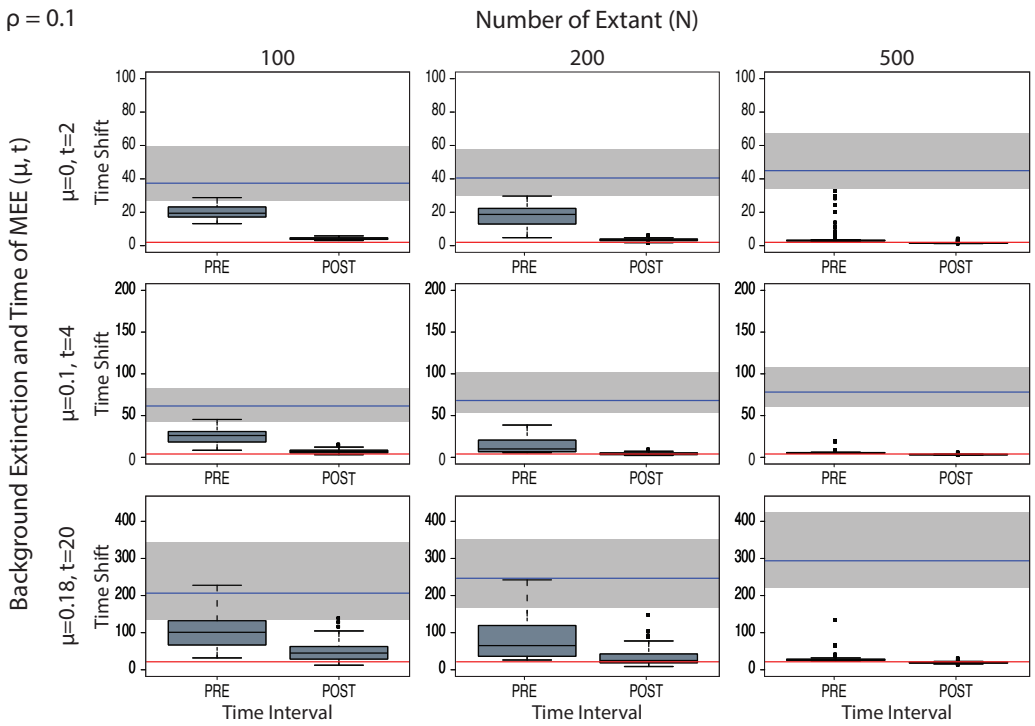


Figure S6: Location of MEEs (as accurate estimation of time intervals for rate changes in μ that contain the MEE) for Model C under varying levels of background extinction and N (number of extant taxa). All other conventions follow Figure 4.

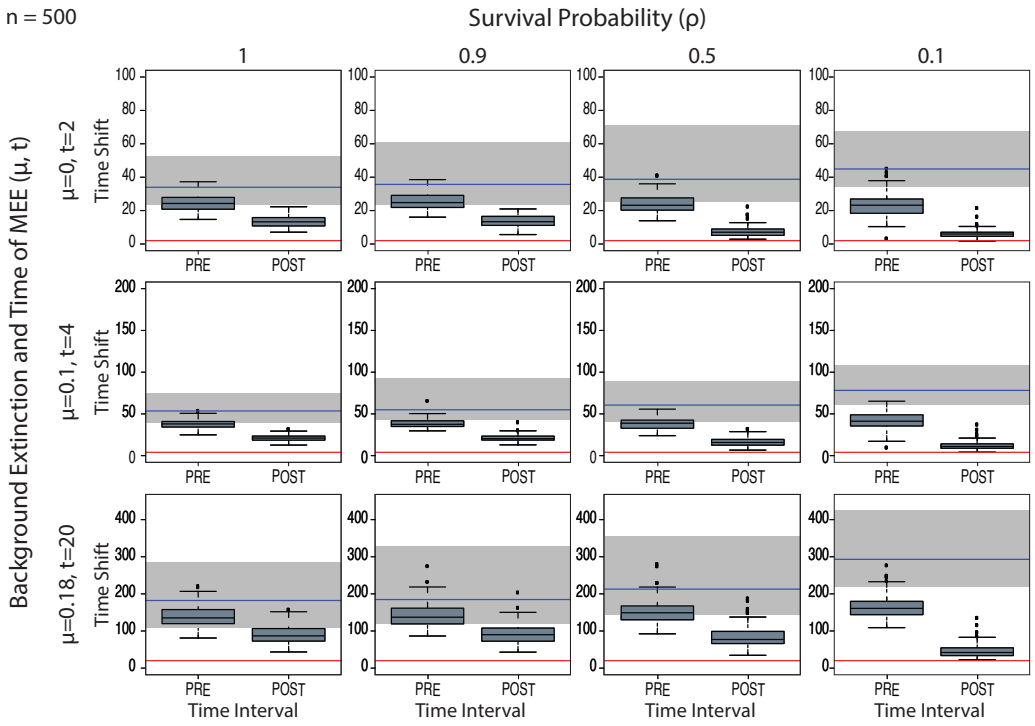


Figure S7: Location of MEEs (as accurate estimation of time intervals for rate changes in μ that contain the MEE) for Model A under varying levels of background extinction and mass extinction survival probability, ρ. All other conventions follow Figure 4.

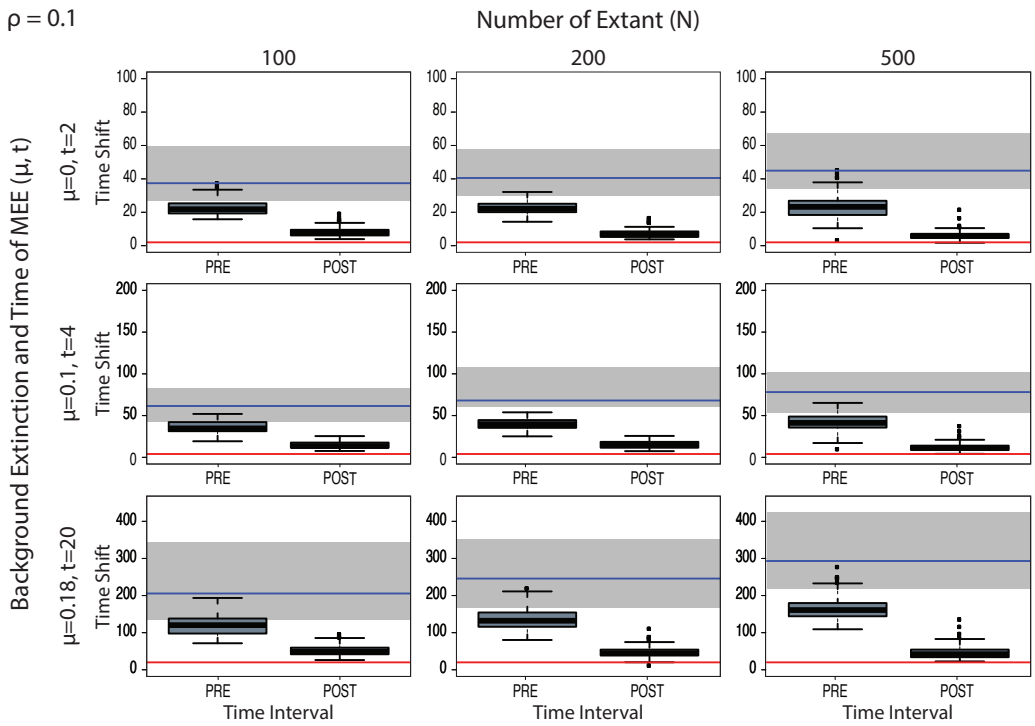


Figure S8: Location of MEEs (as accurate estimation of time intervals for rate changes in μ that contain the MEE) for Model A under varying levels of background extinction and number of extant taxa (N). All other conventions follow Figure 4.

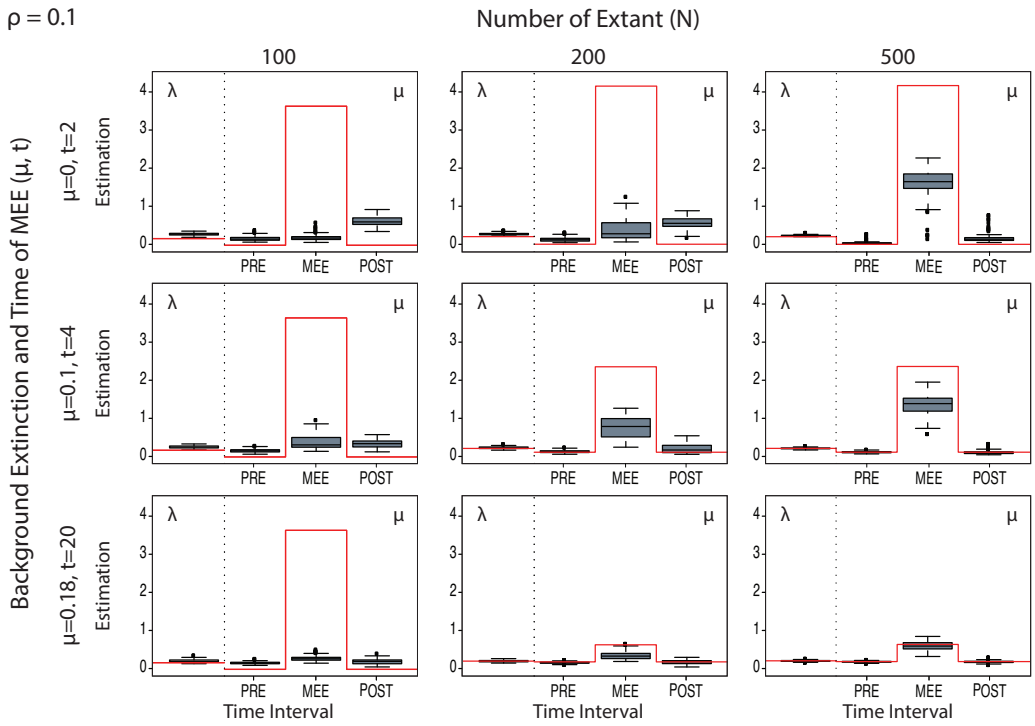


Figure S9: Estimation of changes in magnitude of μ across time intervals for Model C under different values of background extinction and number of extant taxa (N). All other conventions follow Figure 5.

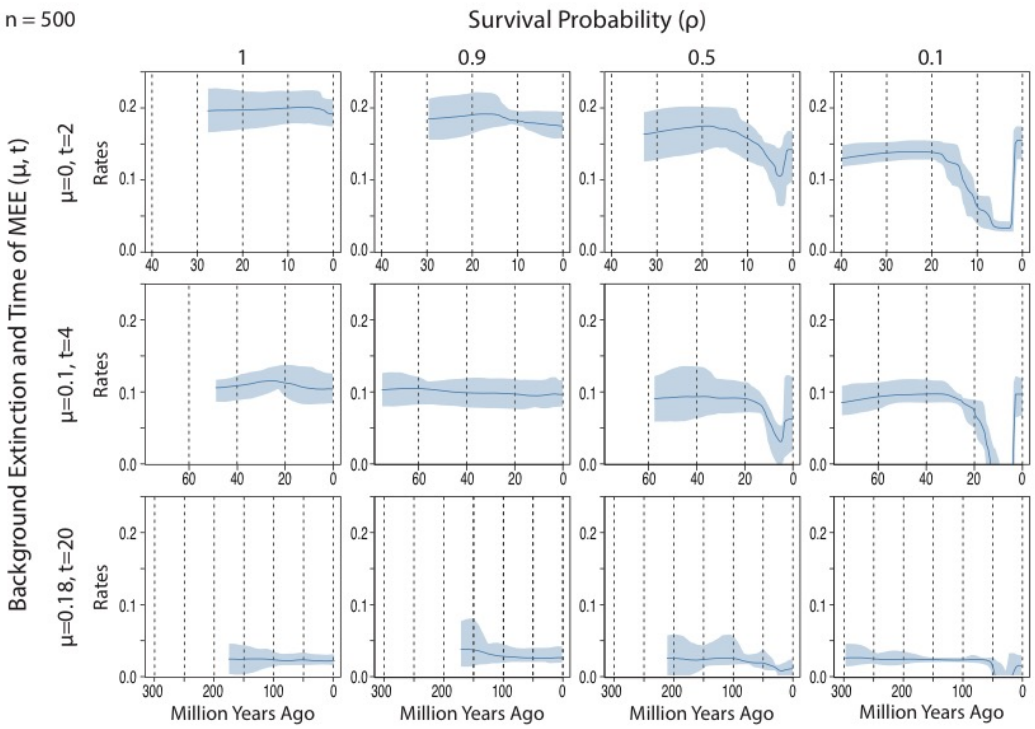


Figure S10 Detection of MEEs through interoperating changes in the magnitude of the diversification rate (diversification = λ - μ) in CoMET under varying levels of μ and mass extinction survival probability, ρ.

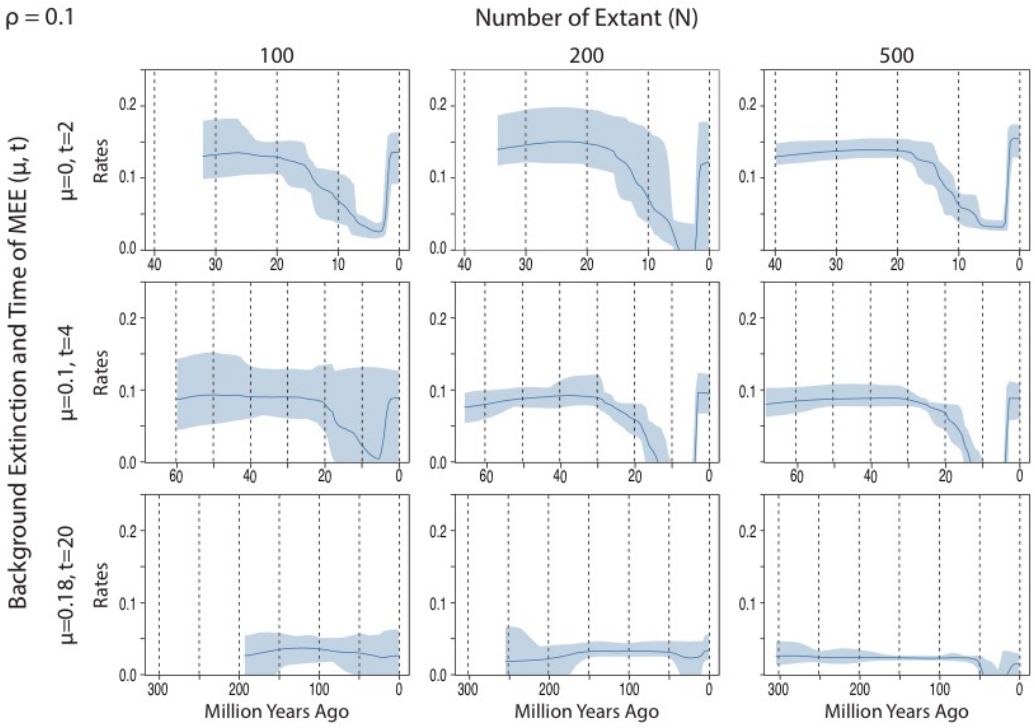


Figure S11 Detection of MEEs through interoperating changes in the magnitude of the diversification rate (diversification = λ - μ) in CoMET under different values of background extinction and number of extant taxa (N).

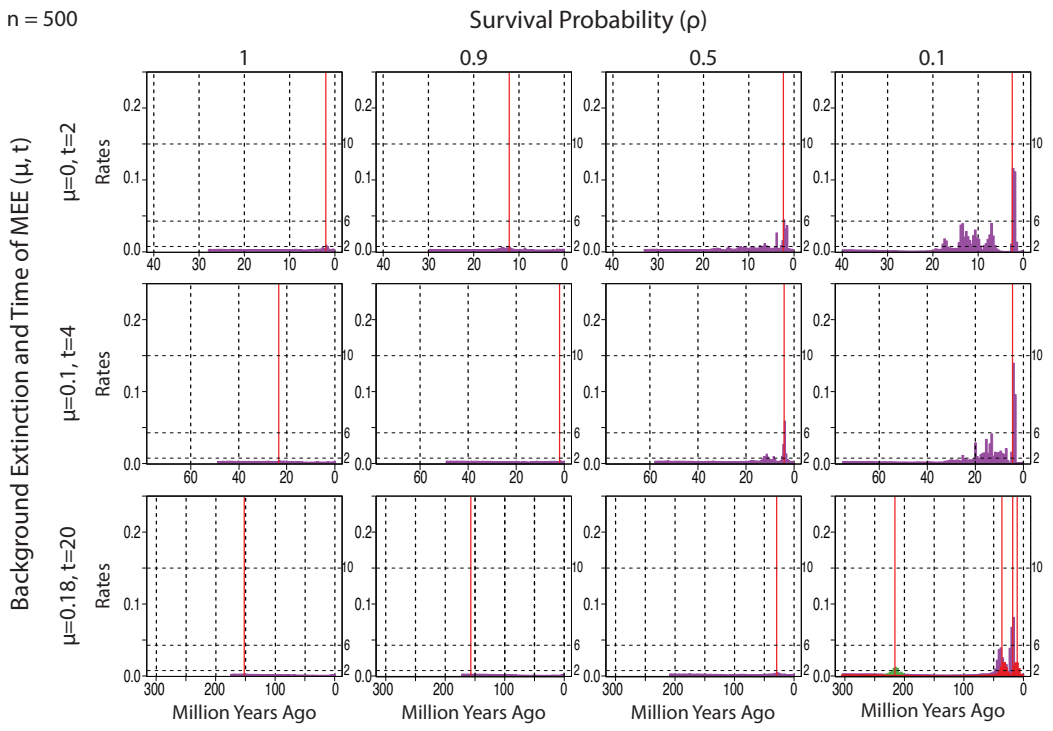


Figure S12 Location of MEEs indicated in the rate changes times in λ (purple) and/or μ (red) and/or *mass extinction time* (green) if returned by CoMET under varying levels of background extinction and mass extinction survival probability, ρ.

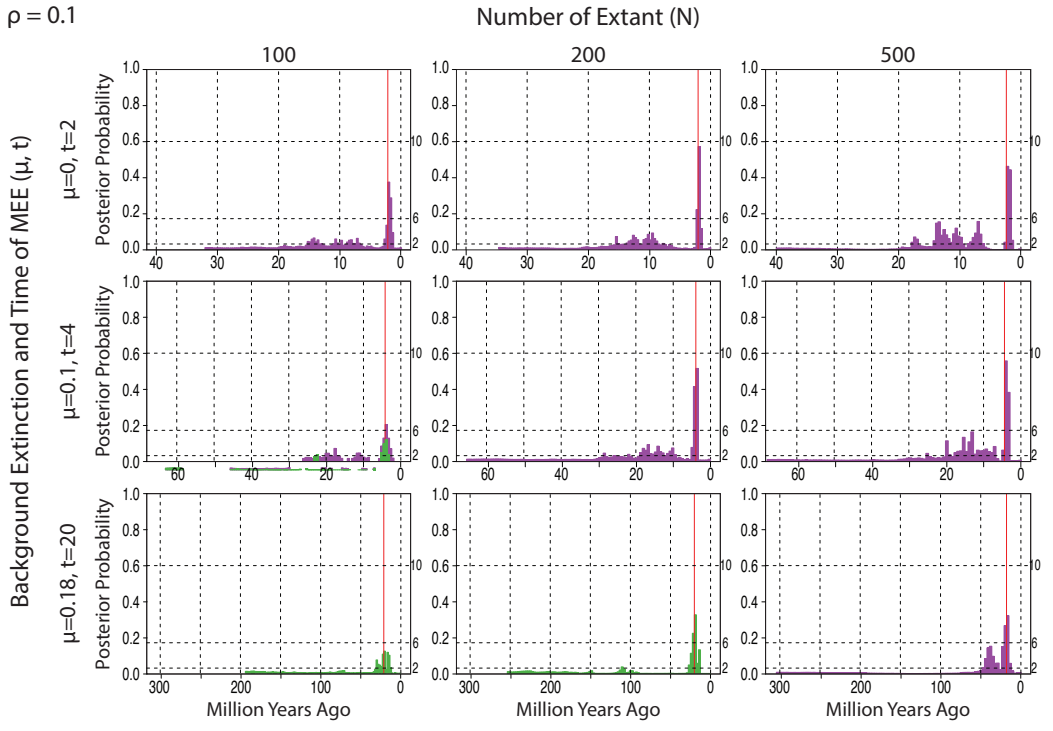


Figure S13 Location of MEEs indicated in the rate changes times in λ (purple) and/or μ (red) and/or *mass extinction time* (green) if returned by CoMET under different values of background extinction and number of extant taxa (N).

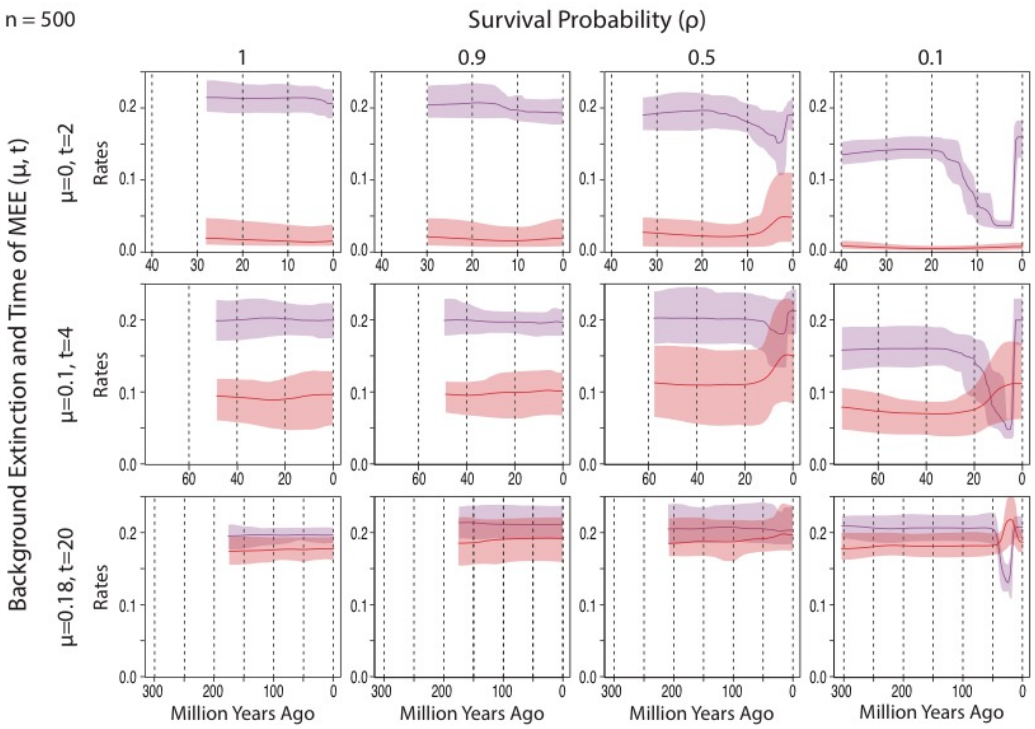


Figure S14 Estimation of changes in magnitude of μ and λ across time intervals for CoMET under different values of background extinction and MEE survival probability, ρ.

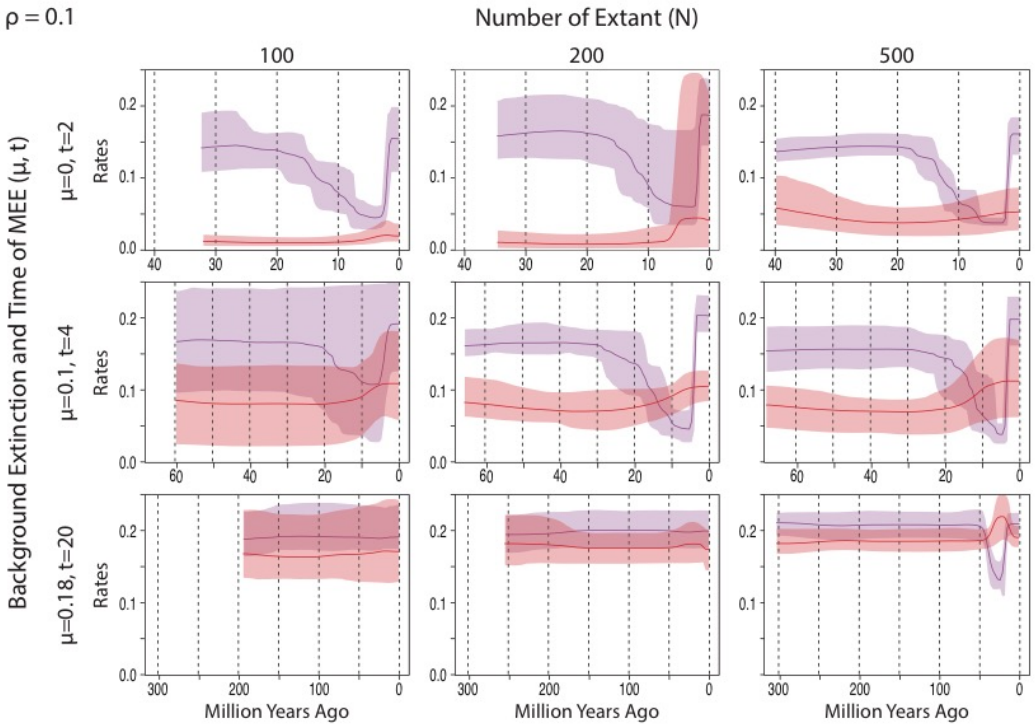


Figure S15 Estimation of changes in magnitude of μ and λ across time intervals for CoMET under different values of background extinction and number of extant taxa (N).

Supplementary Tables

Table S1: Percentage of simulated phylogenies that converged in BEAST2 for each of the three model settings (A, B, and C) of the birth-death skyline model parameter (BDSKY).

μ	**ρ**	**N**	***t***	**Scenario**	**Model A**	**Model B**	**Model C**
					% converged	**% converged**	**% converged**
0	0.1	500	2	1, 2, 3	92	100	100
0.1	0.1	500	4	1, 2, 3	98	100	100
0.18	0.1	500	20	1, 2, 3	97	100	100
0	0.5	500	2	2	100	100	100
0.1	0.5	500	4	2	100	100	100
0.18	0.5	500	20	2	100	100	100
0	0.9	500	2	2	100	100	100
0.1	0.9	500	4	2	100	100	100
0.18	0.9	500	20	2	100	100	100
0	0.1	100	2	3	100	100	99
0.1	0.1	100	4	3	100	100	100
0.18	0.1	100	20	3	100	100	100
0	0.1	200	2	3	88	100	89
0.1	0.1	200	4	3	86	100	96
0.18	0.1	200	20	3	100	100	100

Table S2: Summary statistics for the Birth-Death Skyline model parameters. Models A to C refer to different BEAST2 settings with varying constraints (see text for a detailed description). Abbreviations: "N" = Number of taxa; "μ" = value of extinction rate in simulations; "ρ" = survival probability in simulations; mass extinction intensity = (1- ρ). Posterior probability estimates and accuracy for the Diversification Rate: "Acc": the mean of the means of the estimated parameters across trees; "Prec": the mean of the width of the 95% High Posterior Density (HPD) credibility interval across trees; "Cov": Coverage, percentage of simulated trees where the 95% HPD credibility interval contained the true parameter value; "HPDn" is defined as the percentage of simulated trees where the 95% HPD for the diversification rate falls entirely below 0. Posterior probability estimates and accuracy for the Rate Shift Times: "Prec"; the width of the 95% HPD of the estimated MEE time interval; "Cover": the percentage of simulations where at least 95% of the estimated MEE time interval contains the true time.

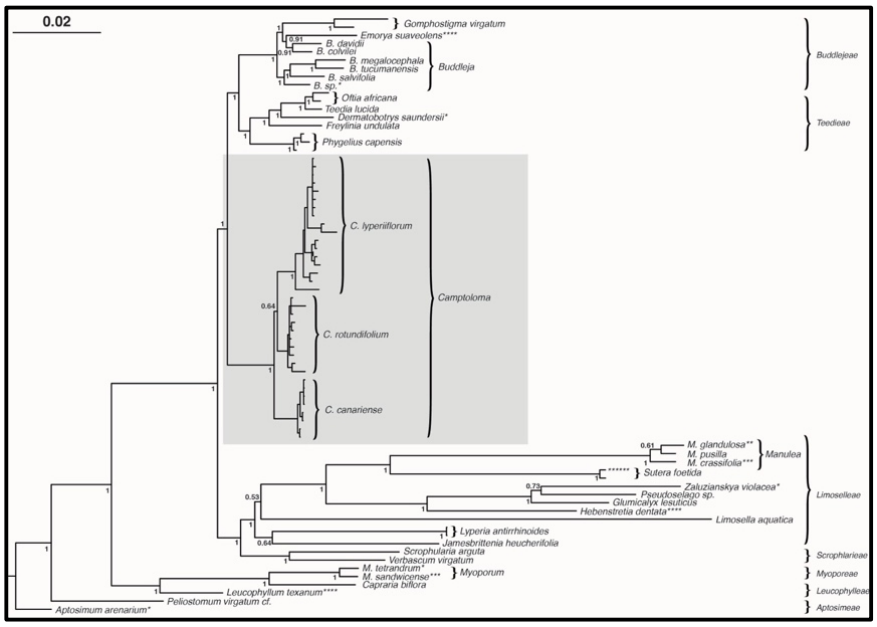
N=500		Survival Probability in Simulations, ρ= 1						ρ= 0.9						ρ= 0.5						ρ= 0.1					
		Diversification Rate				Rate Shift Times		Diversification Rate				Rate Shift Times		Diversification Rate				Rate Shift Times		Diversification Rate				Rate Shift Times	
	Time Interval	Acc	Prec	Cov	HPDn	Prec	Cov	Acc	Prec	Cov	HPDn	Prec	Cov	Acc	Prec	Cov	HPDn	Prec	Cov	Acc	Prec	Cov	HPDn	Prec	Cov
		Model A																							
	pre	0.13	1.06	1	0	0.44	0	0.14	1.04	1	0	0.42	0	0.12	0.94	1	0	0.41	0	0.09	0.55	1	0	0.75	0.03
μ=0.0	MEE	0.07	0.85	1	0	.	.	0.06	0.83	1	0	.	.	0.06	0.78	1	0	.	.	-4.16	0.85	0	0.01	.	.
	post	0.06	0.27	1	0	.	.	0.07	0.25	0.96	0	.	.	0.21	0.39	0.41	0	.	.	0.25	0.42	0.47	0.04	.	.
	pre	0.01	0.93	1	0	0.38	0	0	0.91	1	0	0.38	0	0	0.81	1	0	0.4	0	0.05	0.46	1	0	0.58	0
μ=0.1	MEE	0.02	0.72	1	0	.	.	0.03	0.71	1	0	.	.	0.05	0.64	1	0	.	.	-2.04	0.65	0	0	.	.
	post	0.01	0.2	1	0	.	.	0.01	0.21	1	0	.	.	0.05	0.26	0.99	0	.	.	0.06	0.26	0.97	0	.	.
	pre	-0.11	0.78	1	0	0.43	0	-0.1	0.75	1	0	0.42	0	-0.09	0.69	1	0	0.38	0	-0.05	0.55	1	0	0.41	0
μ=0.18	MEE	0.04	0.65	1	0	.	.	0.02	0.65	1	0	.	.	0.02	0.56	1	0	.	.	-0.37	0.52	0.37	0	.	.
	post	0	0.12	1	0	.	.	0	0.11	1	0	.	.	0	0.12	1	0	.	.	0	0.09	0.98	0	.	.
		Model B																							
	pre	0.21	1.28	1	0	.	.	0.21	1.28	1	0	.	.	0.2	1.29	1	0	.	.	0.2	1.28	1	0	.	.
μ=0.0	MEE	0.21	1.46	1	0	.	.	0.2	1.47	1	0	.	.	0.21	1.46	1	0	.	.	0.21	1.46	1	0	.	.
	post	0.01	0.05	0.9	0	.	.	0.02	0.05	0.56	0	.	.	0.06	0.04	0	0	.	.	0.11	0.02	0	0	.	.
	pre	0.06	1.09	1	0	.	.	0.05	1.11	1	0	.	.	0.01	1.11	1	0	.	.	0.01	1.14	1	0	.	.
μ=0.1	MEE	0.09	1.46	1	0	.	.	0.1	1.44	1	0	.	.	0.08	1.46	1	0	.	.	0.06	1.49	1	0	.	.
	post	0	0.05	0.97	0	.	.	0	0.04	0.93	0	.	.	0.03	0.03	0.05	0	.	.	0.07	0.02	0	0	.	.
	pre	-0.07	0.35	0.95	0	.	.	-0.08	0.34	0.94	0	.	.	-0.1	0.34	0.9	0	.	.	-0.13	0.35	0.76	0	.	.
μ=0.18	MEE	0.05	1.28	1	0	.	.	0.03	1.28	0.98	0	.	.	0.03	1.3	1	0	.	.	0	1.35	1	0	.	.
	post	0	0.02	0.94	0	.	.	0	0.02	0.93	0	.	.	0.01	0.01	0.42	0	.	.	0.02	0.01	0	0	.	.
		Model C																							
	pre	0.17	0.64	1	0	32.21	1	0.17	0.63	1	0	34	0.82	0.12	0.54	0.99	0	35.21	0.97	0.01	0.09	0.98	0	6.69	1
μ=0.0	MEE	0.09	0.45	1	0	.	.	0.09	0.43	1	0	.	.	0.06	0.41	1	0	.	.	-2.96	1.73	0	0.84	.	.
	post	0.06	0.25	1	0	.	.	0.07	0.24	0.95	0	.	.	0.23	0.34	0.13	0.03	.	.	0.15	0.37	0.93	0.01	.	.
	pre	0.05	0.5	1	0	51.01	1	0.05	0.48	1	0	52.08	0.93	0.02	0.34	1	0	53.93	0.92	0	0.03	0.95	0	4.71	1
μ=0.1	MEE	0.03	0.36	1	0	.	.	0.03	0.37	1	0	.	.	0.09	0.51	0.99	0.01	.	.	-1.01	1.41	0.26	0.97	.	.
	post	0.01	0.16	1	0	.	.	0.02	0.17	0.99	0	.	.	0.07	0.21	0.76	0	.	.	0	0.19	1	0	.	.
	pre	-0.04	0.35	1	0	168.5	0.81	-0.05	0.34	0.99	0	172.7	0.86	-0.04	0.3	1	0	196.9	0.86	0	0.02	0.94	0	25.52	1
μ=0.18	MEE	0.02	0.44	1	0	.	.	0.01	0.42	1	0	.	.	0.01	0.37	1	0	.	.	-0.04	0.76	0.94	0.93	.	.
	post	0	0.09	1	0	.	.	0	0.09	1	0	.	.	0.01	0.09	0.96	0	.	.	0	0.06	0.96	0	.	.

Code

All code from this study is deposited in the public repository Dryad, under number https://doi.org/10.5061/dryad.qv10c62.

CHAPTER 2: APPENDIX

Combining Bayesian biogeographic inference and phylogenetically-informed niche models to reconstruct the role of ancient climate change in depauperate lineages



Extended Material and Methods

This section provides details on Material and Methods not included in the main article.

DNA Sequencing

PCR amplifications were performed using an Eppendorf Mastercycler Epgradient S (Westbury, NY). With the old herbarium material, we experimented with longer extraction incubation times (Villaverde et al. 2018), while for some samples exhibiting more than one band amplification of PCR product from gel was needed.

We selected seven non-coding cpDNA regions exhibiting moderate or high levels of genetic variation. The intergenic spacers where: *trn*L-*trn*F, *trn*S-*trn*G, *rpl*32-*ndh*F, *psb*J-*pet*A, *pet*B-*pet*D, *trn*T-*trn*L and the *rps16* intron. The spacer *rpl*32-*trn*L (*trn*L(UAG)-*rpl*32-F) (Shaw et al. 2007) was originally included but the alignments returned were of poor quality, so this spacer was eventually excluded. Most samples were amplified using universal standard primers (Taberlet et al. 1991; Hamilton 1999; Cronn et al. 2002; Löhne et al. 2005; Shaw et al. 2005, 2007). Some old herbarium samples were difficult to amplify, and for these we designed new custom primers using the software "primer3" (Koressaar et al. (2007); Untergasser et al. (2012) in Geneious Pro v5.7.6 (www.geneious.com). We also sequenced the multicopy nuclear marker ITS using universal primers (White et al. 1990), with the difficult samples being amplified using plant specific primer pairs ITS p4-p5, ITS p2-p5, and ITS p3-p4 (Cheng et al. 2015). PCR products were sequenced in Macrogen Inc., the Netherlands (www.macrogen.com). Table S2 gives primer sequences and specific PCR protocols. Table S1 provides (www.ncbi.nlm.nih.gov/genbank) to compensate for failed amplified sequences (Table S1).

The best-fit DNA substitution model for each marker was selected using jModelTest v.2.1.10 (Darriba et al. 2012). These were GTR+I+G for ITS; TVM+G (*rpl*32-*ndh*F); TIM3+G (*pet*B-*pet*D), GTR+G (*psb*J-*pet*A), TVM+G (*rps*16), *trn*L-*trn*F (TVM+G), TrN+G (*trn*S-*trn*G), TVM+G (*trnT-trnL*). When models selected by jModelTest were not implemented in MrBayes, we used the next more parameterised model, GTR+G for chloroplast markers and GTR+I+G for ITS. The same models were used for the cpDNA and ITS partitions, respectively, in the concatenate dataset.

Dating and Biogeographic Inference

For the non-nested analysis, we used a relaxed uncorrelated lognormal clock (UCLD), which was linked for all nucleotide partitions, with a random starting tree and a birth-death prior with the "incomplete Sampling" option selected (Stadler 2009). All priors were set to default except for *ucld.sd* (Lognormal[mean=R0.8, standard deviation=0.1] , where "R" means "in real space") and *ucld.mean* (Lognormal[R0.003, 0.2]). Mixing and convergence were assessed in Tracer (each analysis was run until ESS reached at least 200).

For the phylogeographic inference with DTA, identically distributed, independent gamma priors (alpha=1) were used for transition rates, whereas the overall migration rate was modelled using the default continuous-time Markov Chain reference prior (CTMC, Ferreira et al. 2008). Independent substitution models were implemented for cpDNA (GTR+G) and nuclear (GTR+I) markers; a coalescent constant-size population model was used for the tree growth prior and a lognormal uncorrelated relaxed clock (*ucld.sd* and *ucld.mean* as above) for the molecular clock prior.

For the BASTA analysis, we implemented an unlinked site model partitioned into cpDNA DNA (GTR+G), ITS (GTR+I+G) and location (the "migration-mutation Volz model" with symmetric transition rates and equal population sizes across areas i.e. a social network with BSSVS inferred; a relaxed UCLD clock model (with "estimation" selected and rate = 0.002 given, Drummond et al. 2006) was used for the nucleotide partitions with ucld.sd and ucld.mean as above, and a strict clock model for the biogeographic character. Priors for the location rates where calculated using the Discrete Rate Prior Generator (multivariate Gamma Prior with reversible priors that uses normalised inverse distances obtained from the population coordinates). All other priors were set as default. As in DTA, the topology and molecular branch lengths were fixed to the nested-dating MCC tree. Mixing and convergence were assessed in Tracer (each analysis was run until ESS reached at least 200).

Supplementary References

Cheng T., Xu C., Lei L., Changhao L., Zhang Y., Zhou S. 2015. Barcoding the kingdom Plantae: new PCR primer for ITS regions of plants with improved universality and specificity. Molecular Ecology Resources. 16:10.1111/1755-0998.12438.

Cronn R.C., Small R.L., Haselkorn T., Wendel J.F. 2002. Rapid diversification of the cotton genus (Gossypium: Malvaceae) revealed by analysis of sixteen nuclear and chloroplast genes. American Journal of Botany. 89:707-725.

Drummond A.J., Ho S.Y.W., Phillips M.J., Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biology. 4:e88.

Ferreira M.A.R., Suchard M.A. 2008. Bayesian analysis of elapsed times in continuous-time Markov chains. Canadian Journal of Statistics. 36:355-368.

Hamilton M.B. 1999. Four primers pairs for the amplification of chloroplast intergenic regions

with intraspecific variation. Molecular Ecology. 8:521-523.

Koressaar T., Remm M. 2007. Enhancements and modifications of primer design program Primer3. Bioinformatics. 23:1289-1291.

Löhne C., Borsch T. 2005. Molecular evolution and phylogenetic utility of the petD group II intron: a case study in basal angiosperms. Molecular Biology and Evolution. 22:317-332.

Shaw J., Lickey E.B., Beck J.T., Farmer S.B., Liu W., Miller J., Siripun K.C., Winder C.T., Schilling E.E., Small R.L. 2005. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. American Journal of Botany. 92:142-166.

Shaw J., Lickey E.B., Schilling E.E., Small R.L. 2007. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogentic studies in Angiosperms: the tortoise and the hare III. American Journal of Botany. 94:275-288.

Taberlet P., Gielly L., Pautou G., Bouvet J. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Molecular Biology. 17:1105-1109.

Untergasser A., Cutcutache I., Koressaar T., Ye J., Faircloth B.C., Remm M., Rozen S.G. 2012. Primer3–new capabilities and interfaces. Nucleic Acids Research 40:e115.

White T.J., Bruns T., Lee S., Taylor J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. PCR Protocols: A Guide to Methods and Applications. 18:315-322.

Supplementary Figure

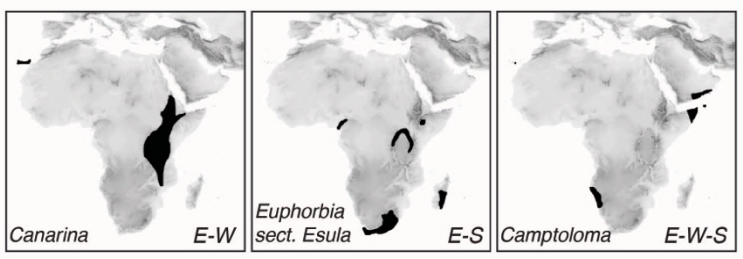


Figure S1. Rand Flora disjunct distributions with representative lineages.

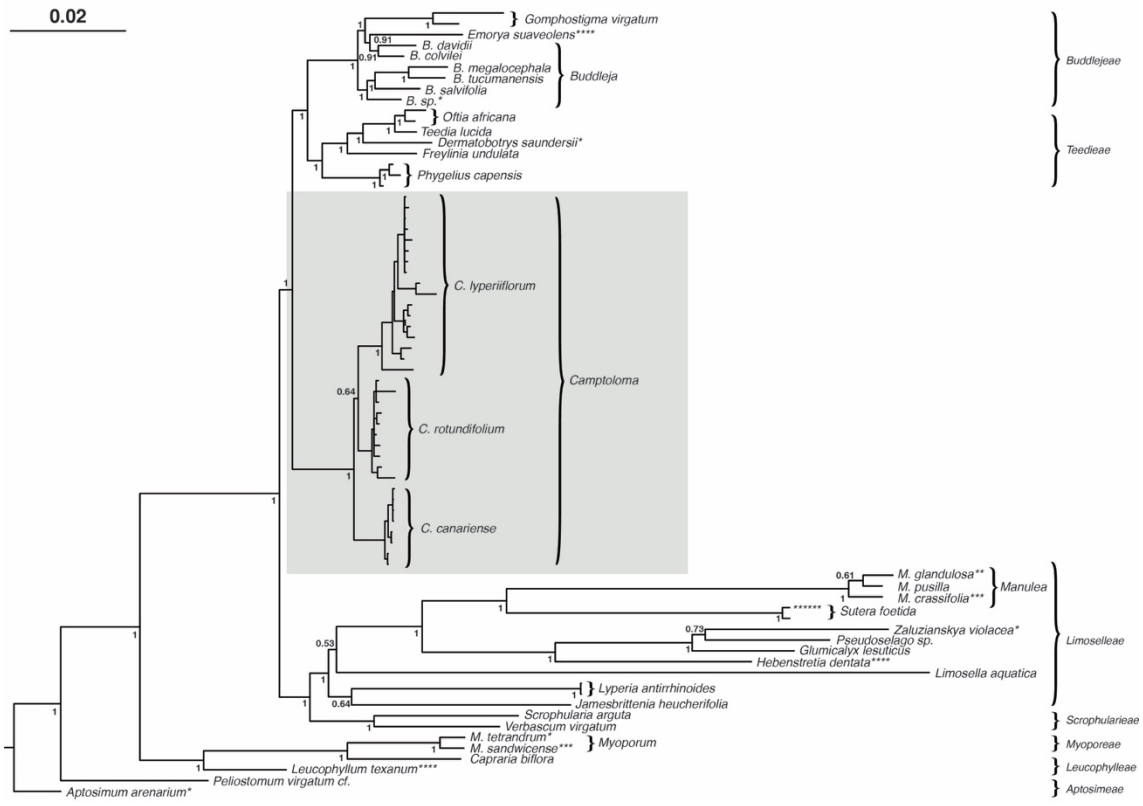


Figure S2. Phylogeny of genus *Camptoloma* and representative outgroups within Scrophulariaceae. Bayesian majority-rule consensus tree obtained in MrBayes using the concatenate "outgroup dataset" of nuclear (ITS) and chloroplast DNA *(trn*L-*trn*F*, trn*S-*trn*G*, rpl*32-*ndh*F*, psb*J-*pet*A*, pet*B-*pet*D*, trn*T-*trn*L and *rps*16 *intron*) markers. Numbers in branches indicate clade posterior probability values (PP);

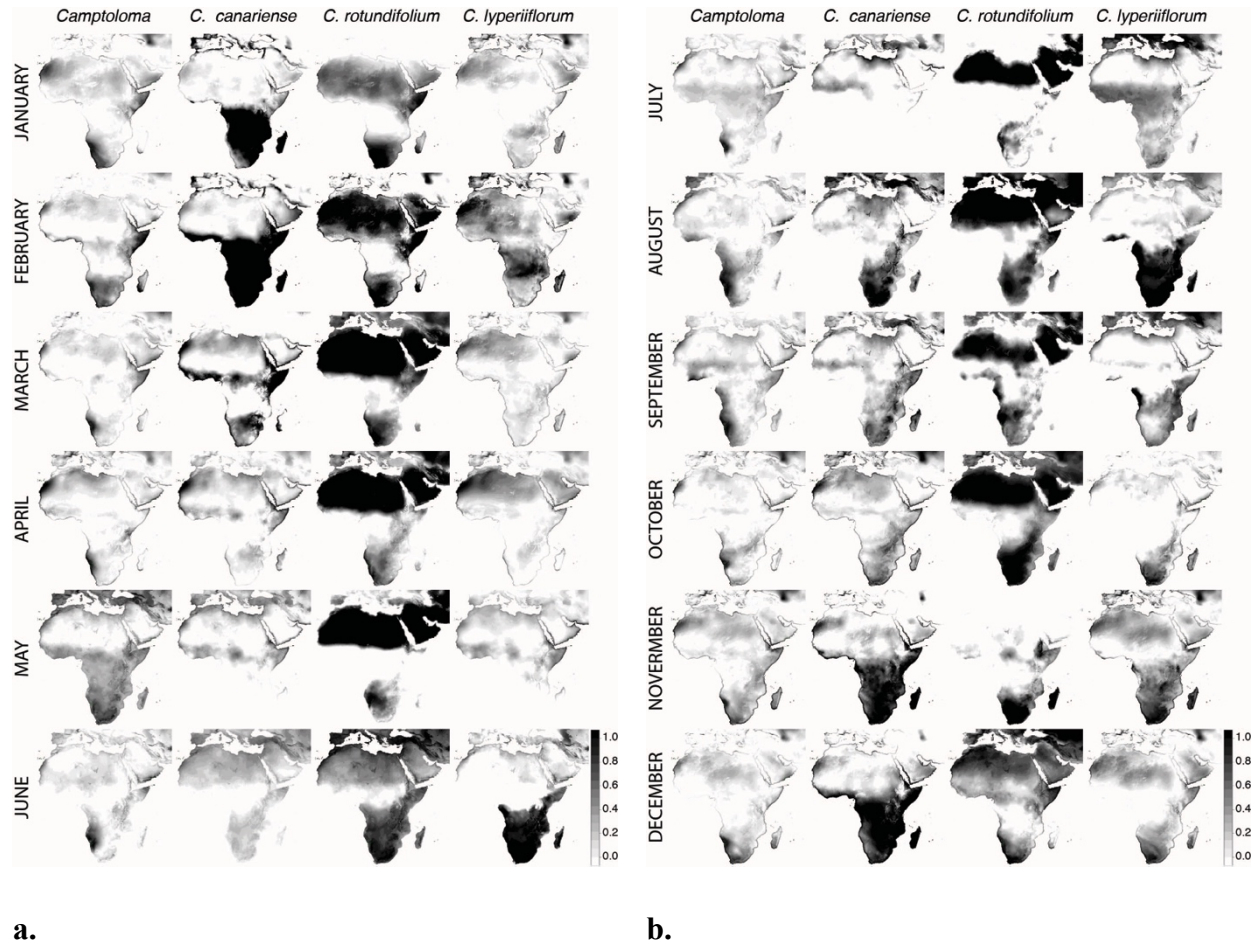
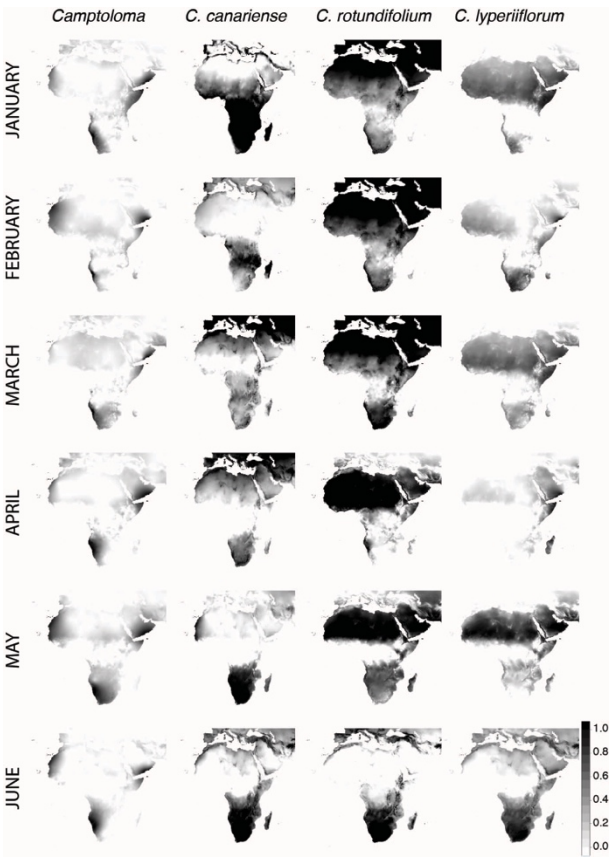
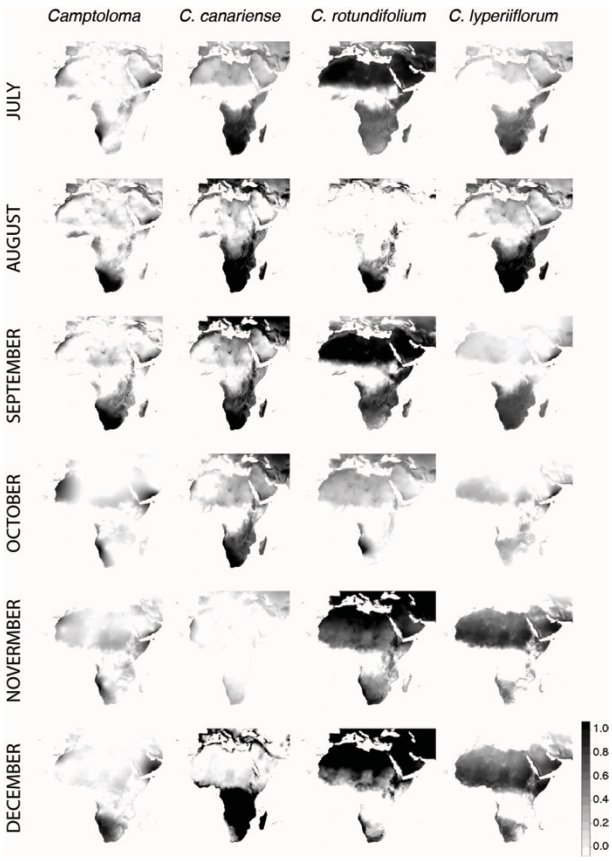


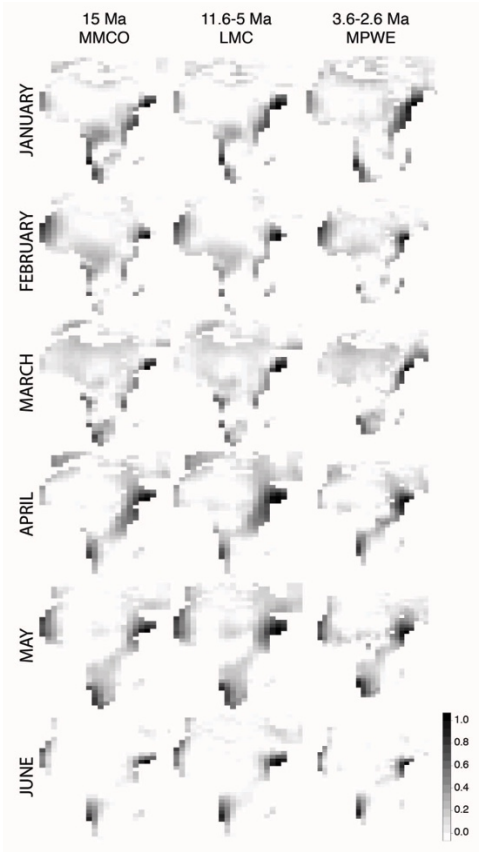
Figure S3. The raw values of present-day ENMs of the full model for: **a)** each *Camptoloma* species, **b)** the genus, across the months of the present-day year before being run through the truncated threshold of 0.7. The grey scale indicates patch suitability with black indicating high suitability and white low.


Camptoloma
C. canariense
C. rotundifolium
C. lyperiiflorum
JANUARY
FEBRUARY
MARCH
APRIL
MAY
JUNE
1.0
0.8
0.6
0.4
0.2
0.0

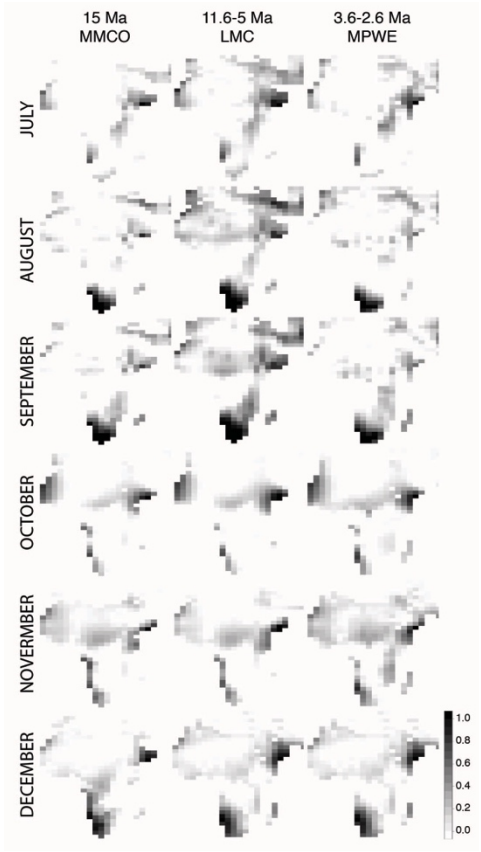
a.


Camptoloma
C. canariense
C. rotundifolium
C. lyperiiflorum
JULY
AUGUST
SEPTEMBER
OCTOBER
NOVERMBER
DECEMBER
1.0
0.8
0.6
0.4
0.2
0.0

b.

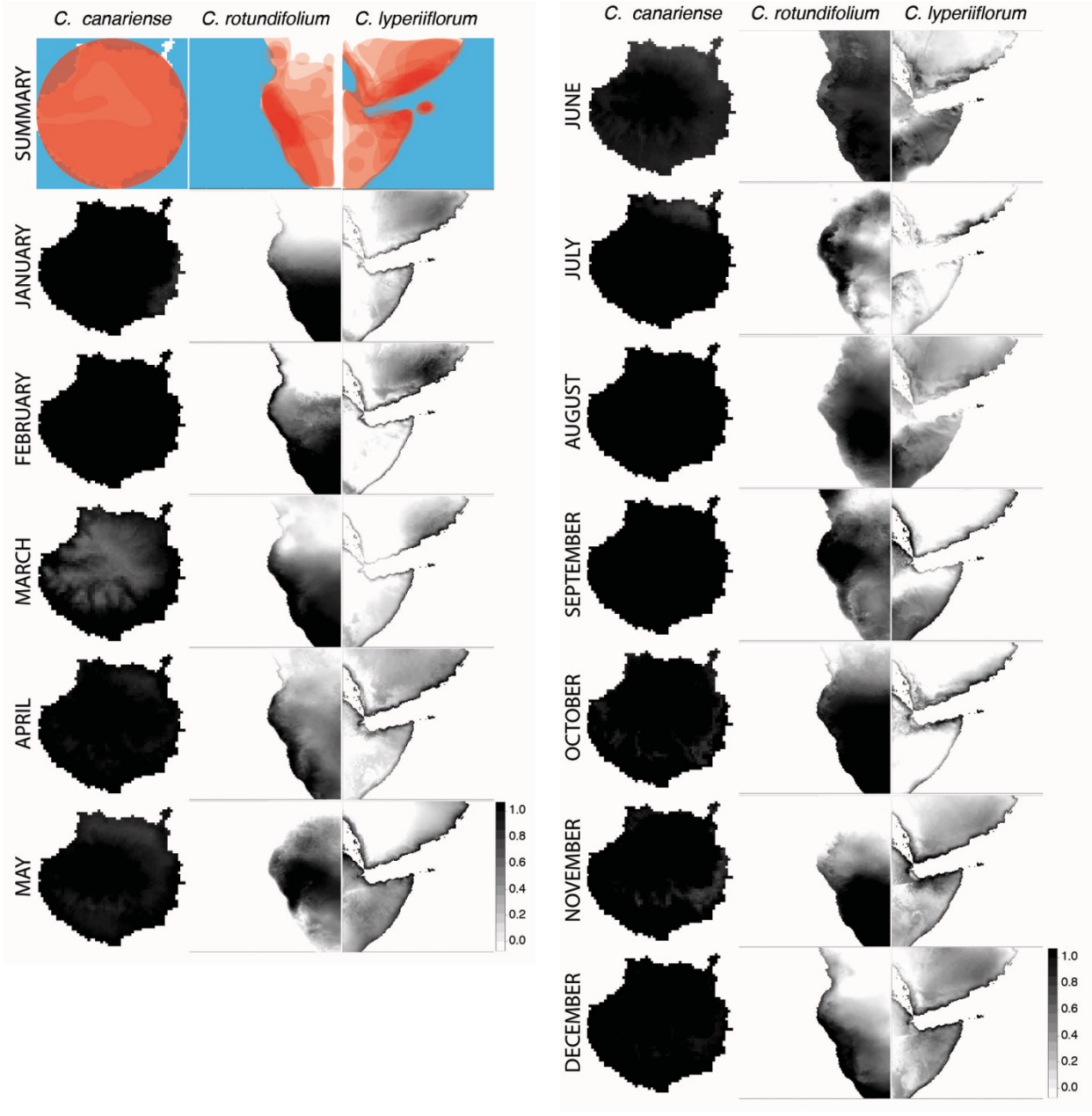

15 Ma
MMCO
11.6-5 Ma
LMC
3.6-2.6 Ma
MPWE
JANUARY
FEBRUARY
MARCH
APRIL
MAY
JUNE
1.0
0.8
0.6
0.4
0.2
0.0

c.


15 Ma
MMCO
11.6-5 Ma
LMC
3.6-2.6 Ma
MPWE
JULY
AUGUST
SEPTEMBER
OCTOBER
NOVERMBER
DECEMBER
1.0
0.8
0.6
0.4
0.2
0.0

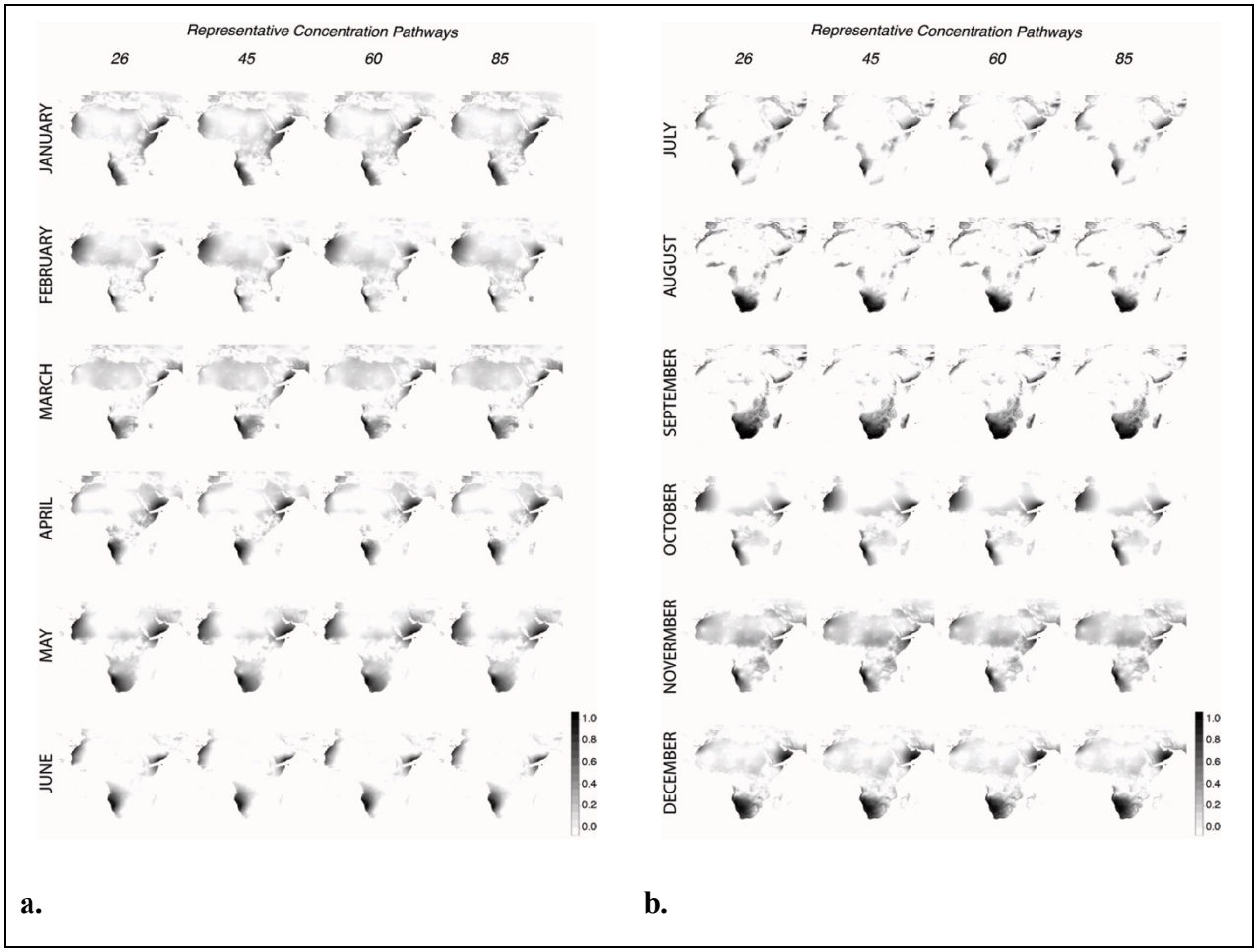
d.

Figure S4. The raw values of present-day potential habitat suitability for the reduced model for: each *Camptoloma* species and for the genus across the months of the present-day year before being run through the truncated threshold of 0.7 **(a, b)**, and for the genus only back through time before the values were run through a truncated threshold of 0.1 **(c, d)**. The grey scale indicates patch suitability with black indicating high suitability and white low.



a. **b.**

Figure S5. (a) The raw values of present day potential habitat suitability for the full model for each *Camptoloma* species in their known present day location across the months of the present-day year before being run through the truncated threshold of 0.7. The grey scale indicates patch suitability with black indicating high suitability and white low. **(b)** The summary represents the clustering of habitat suitable area post-truncated threshold for a combined year.


Representative Concentration Pathways
26
45
60
85
JANUARY
FEBRUARY
MARCH
APRIL
MAY
JUNE
1.0
0.8
0.6
0.4
0.2
0.0
Representative Concentration Pathways
26
45
60
85
JULY
AUGUST
SEPTEMBER
OCTOBER
NOVERMBER
DECEMBER
1.0
0.8
0.6
0.4
0.2
0.0

a. **b.**

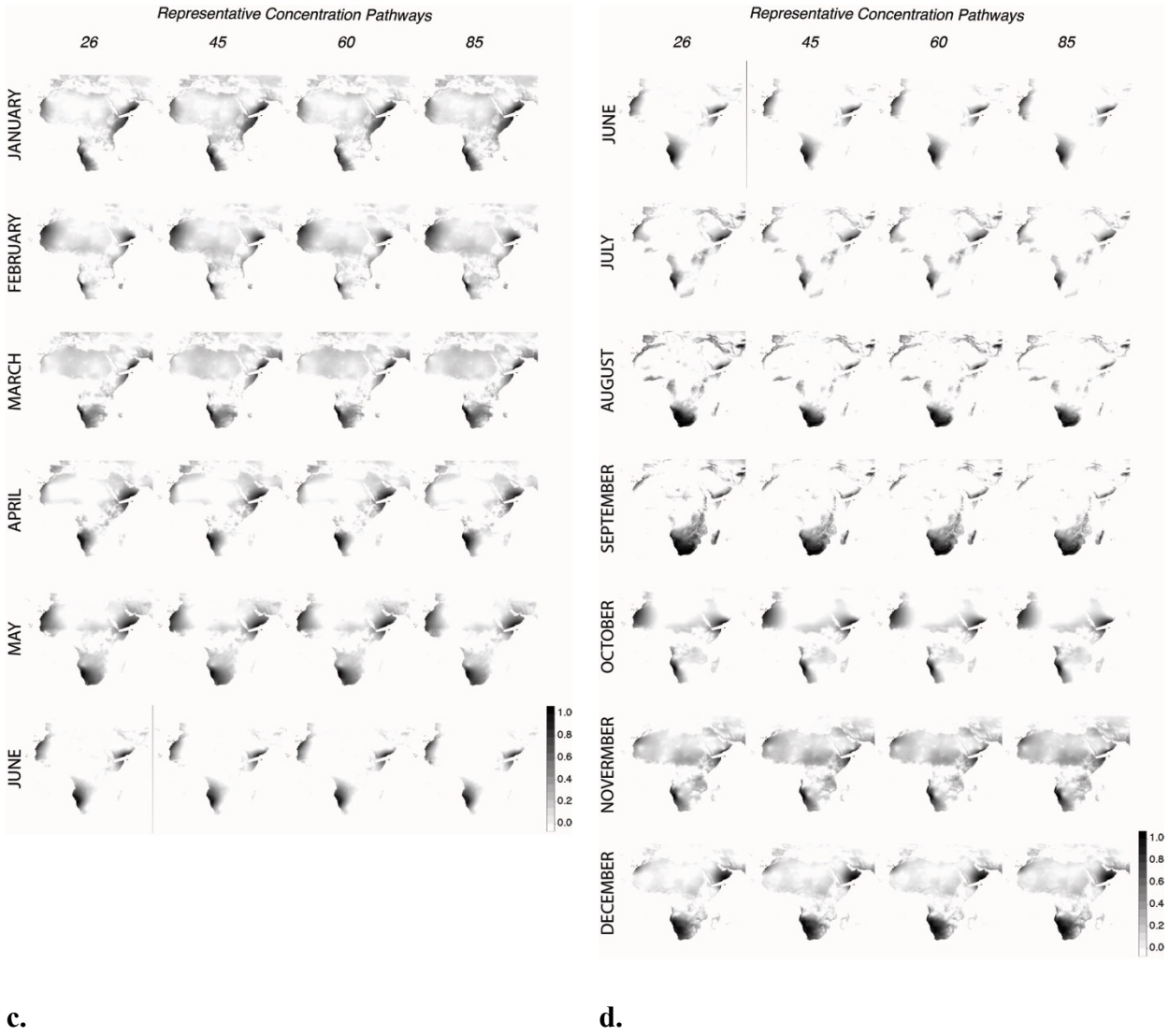


Figure S6. The raw values of forecasted potential habitat suitability for the reduced model for each *Camptoloma* genus across the months for year intervals 2041-2060 **(a, b)** and 2061-2080 (**c, d)** for the four global warming scenarios before the values were run through the truncation threshold of 0.7. The grey scale indicates patch suitability with black indicating high suitability and white low.

Supplementary Tables

. **Table S1.** Taxon names, location, voucher information, and GenBank accession numbers for all taxa included in this study.

Taxon	Country; Locality	Geographic Coordinates	Collector; Coll. # (Herbarium)	ITS (1A-4)	ndhF (rpl32 R-ndhF)	petB-D (petB 1365F-petD 738R)	psbJ-petA	rps16intron (rps16F-rpsR)	trnL (C-F)	trnS-G (trnS GCU-trnG UCC)	trnT-L (trnA2-trnB)
Aptosimum arenarium	Nambia; Distr. Khomas, Gamsberg pass, road C26	23.2694667 16.3311667	Medina; LM6056 (MA)			–					
Buddleja colvilei	–;	–	M.W. Chase; 37994 (K)								
Buddleja davidii	–;	–	M.F. Fay; MFF179 (K)								
Buddleja megalocephala	–;	–	M.J.M. Christenhusz; 5266 (BM, MO, H)								
Buddleja salvifolia	Spain; Madrid RJB	–	Cleary; 67 (MA)								
Buddleja tucumanensis	South America; Bolivia Cochabamba	-17.2869444 -66.3105556	Aedo; 14514 (MA)								
Buddleja sp.	–; –		Felix Forest; F747 (K)	–		–		–			
Camptoloma canariensis	Spain; Gran Canaria Barranco del Draguillo	27.9471944 -15.4379444	M. Mairal; MM4.1 (RJB)								

Camptoloma canariensis	Spain; Gran Canaria Barranco del Draguillo	27.9471944 -15.4379444	M. Mairal; MM4.2 (RJB)								
Camptoloma canariensis	Spain; Gran Canaria Barranco de Guayadeque	27.9332639 -15.5112806	M. Mairal; MM2.1 (RJB)								
Camptoloma canariensis	Spain; Gran Canaria Barranco de Guayadeque	27.9332639 -15.5112806	M. Mairal; MM2.2 (RJB)								
Camptoloma canariensis	Spain; Gran Canaria Degollada de Tasarte	27.8193139 -15.5792778	M. Mairal; MM67 (RJB)								
Camptoloma canariensis	Spain; Gran Canaria Roque Bentayga	27.9909 -15.642025	M. Mairal; MM82 (RJB)								
Camptoloma canariensis	Spain; Gran Canaria Degollada de Tasarte	–	M. Mairal; MM69 (RJB)								
Camptoloma canariensis	Spain; Gran Canaria Roque Bentayga	–	M. Mairal; MM84 (RJB)								
Camptoloma lyperiiflorum	Oman; Oman Dhofar (ufār) Governorate	17.3266667 55.2858333	M. Raffaelli, M. Tardelli, M. Masi; 17-346 (FT)								
Camptoloma lyperiiflorum	Oman; Shalim and the Hallaniyat Islands Province, Kuria Muria Islands	–	Anthony G. Miller; 2755 (E & K)								
Camptoloma lyperiiflorum	Oman; Shalim and the Hallaniyat Islands	17.5166667 56.0666667	M.D. Gallagher; 6252/4 (E)				–				

	Province, Kuria Muria Islands										
Camptoloma lyperiiflorum	Oman; Shalim and the Hallaniyat Islands Province, Kuria Muria Islands	17.5333333 56.0500000	Michael D. Gallagher; 6602/2 (E)				–		–		
Camptoloma lyperiiflorum	Oman; Shalim and the Hallaniyat Islands Province, Kuria Muria Islands	17.5000000 56.1333333	Ian McLeish; 1576 (E)			–			–		
Camptoloma lyperiiflorum	Yemen; 6 km SE of Qalansiyah	12.635528 53.517496	A.G. Miller, L. Guarino, N. Obadi, M. Hassan & N. Mohammed; M.8391 (E & K)								
Camptoloma lyperiiflorum	Oman; , Kuria Muria Islands	17.5333333 56.0333333	M.D. Gallagher; 6252/21 (E)			–	–		–		
Camptoloma lyperiiflorum	Oman; , Kuria Muria Islands	17.5166667 56.0500000	M.D. Gallagher; 6252/11 (E)				–		–		
Camptoloma lyperiiflorum	Yemen; Mukalla Quaiti State	14.53711 49.15239	O. Kerfoot; 3011 (K)								
Camptoloma lyperiiflorum	Yemen; Raidat al Maarah, Kathiri	16.928517 49.369941	O. Kerfoot; 3075 (K)							–	
Camptoloma lyperiiflorum	Oman; Halaniyah Island, aafal beach	17.5166667 58.1500000	A. Patzelt & A. Al-Farsi; HA20 (K)							–	

Camptoloma lyperiiflorum	Somalia; Nugaal Region	08.0500000 49.8000000	Lavranos & Carter; 23496 (K)								–
Camptoloma lyperiiflorum	Somalia;	7.972212 49.833986	P.R.O. Bally & R. Melville; 15537 (K)							–	–
Camptoloma lyperiiflorum	Yemen; Al-Mahra	16.5033333 52.5850556	N. Kilian; YP859 (C)								
Camptoloma lyperiiflorum	Yemen; Al-Mahra	16.6666667 53.0666667	Peter Hein; 6692 (C)								
Camptoloma rotundifolium	Namibia; Erongo Region	-21.134094 14.662936	J. Lundgren & Bertil Nordenstam; 908 (E)								
Camptoloma rotundifolium	South west Africa; Kunene Region	-17.4756528 12.0159667	F.R. Robinson; RK25 (PRE)			–		–	–	–	–
Camptoloma rotundifolium	South west Africa; namibia, orupembe/kaokoveld	-19.1166667 13.7166667	C.J. Ward; 12051 (PRE)								
Camptoloma rotundifolium	South west Africa; namibia, orupembe/kaokoveld	–	R. Story; 5763 (PRE)				only 2 half				
Camptoloma rotundifolium	South west Africa; –	–	Vahrmeijer; 2574 (PRE)								
Camptoloma rotundifolium	South west Africa; Erongo Region, Brandberg	-21.120165 14.417833	Hardy; 1499 (PRE)			–	only 2 half	–	–	–	

Camptoloma rotundifolium	South west Africa; Erongo Region, Nama Karoo	-20.4166667 14.5166667	Burgoyne; 5184 (PRE)								
Camptoloma rotundifolium	South west Africa; Erongo Region, Nama Karoo	-20.596797 14.373419	W. Giess; 7910 (PRE)								
Camptoloma rotundifolium	South west Africa; Erongo Region, Brandberg	-21.134094 14.662936	W. Giess; 5011 (PRE)			–	only 2 half	–	–	–	–
Camptoloma rotundifolium	Namibia; Kunene Region	-20.596993 14.373441	W. Giess; 7911 (K)								
Camptoloma lyperiiflorum	Yemen; Socotra	12.3500000 53.5666667	M. Thulin & A.N. Gifri; 8572 (K)								
Camptoloma lyperiiflorum	Yemen; Al Mahrah Governorate, Sayhut District	15.2500000 51.3333333	Thulin, Eriksson, Gifri & Långström; 8409 (K)								
Capraria biflora	Panama; San Blas, Nalunega	–	Alonso; 15365JLFA (MA)								
Dermatobotrys saundersii	South Africa;	–	C.D Fraser Jenkins; H/4193/65 (K)			–					
Emorya suaveolens	Mexico; Coahuila	–	F. Lyle Wynd & C. H. Mueller; 154 (K)			–	–	–	–		
Freylinia undulata	–; –	–	P. Goldblatt & L. Porter; 12601 (K)								

Glumicalyx lesuticus	South Africa; Kwazulu-Natal; Sani Pass	-29.5852778 95.2872222	Aedo; 15001 (MA)								
Gomphostigma virgatum	Namibia; –	Grid Ref. 2317 BA Rehoboth	R.D.A. Boylins; BRI B 356 (K)								
Gomphostigma virgatum	South Africa; Transvaal, Pretoria Distract	-25.6166667 28.3666667	Leeuwenberg; 10919 (MA)								
Hebenstretia dentata	Ethiopia; Harerge, Road from Asbe Tefari to Kulubi	9.3833333 41.4000000	Jansen; 2512 (MA)			–	–		AJ296499.1	–	
Jamesbrittenia heucherifolia	Angola; Namibe Prov.	-16.1999722 12.4001111	F. Crawford; 2512 (MA)								
Leucophyllum texanum	Mexico; Casablanca	–	Barkley; FC408 (K)	–			–		JX985277.1	–	–
Limosella aquatica	Spain; Avilia, Rosarito	–	Luceño; 16014 (MA)								
Lyperia antirrhinoides	South Africa; Cape, Malmesbury Distr.: Bothmaskloof Pass	Grid Ref. 3318 BD	J.P.H. Acocks; s.n. (MA)								
Lyperia antirrhinoides	South Africa; Cape, Burned clay slopes near foot of Hex River Pass	Grid Ref. 3319 BC Worcester	P. Goldblatt & J. Manning; 24321 (K)								
Manulea crassifolia	Lesotho; Masere District, Blue Mountains Pass	Grid Ref. 2227 BD	O.M. Hilliard & B.L. Burtt; 9433 (K)			–	–			–	

Manulea glandulosa	South Africa; NW Cape, Gifberg Pass	Grid Ref. 3118 DC Van Rhynsdorp.	L. Hugo; 12097 (K)			–				–	
Manulea pusilla	South Africa; SW Cape, French Hoek Pass	Grid Ref. 3319 CC Worcester	M. Jordaan; 711 (K)								
Myoporum sandwicense	US; Hawaii, Kauai Island, National Botanical Garden	19.750000 -155.250000	H.T. Beck; 7517 (K)				–				
Myoporum tetrandrum	Australia; Esperance District	–	E.N.S. Jackson; 1271 (K)			–	–		–		
Oftia africana	–; –	–	Forest & Manning; 1229 (K)								
Oftia africana	–; –	–	M.W. Chase; 548 (K)								
Peliostomum virgatum-cf	South Africa; Northern Cape, Springbok	–	Buira; 5747 (K)								
Phygelius aff capensis	South Africa; Eastern Cape, Rhodes	-30.7191667 27.9575000	Aedo; WGS84 (MA)								
Phygelius capensis	–; –	–	M.J. Crawley; 15124 (MA)								
Phygelius capensis	–; –	–	Herbarium Kewense Cultivated Plants; MJC436 (K)								

Pseudoselago sp	South Africa; West Cape, Mantagu Pass	-33.9063889 22.4188889	Aedo; n.a. (K)								
Scrophularia arguta	Tunez; Gobernacion de Gafsa	34.4391667 9.2613889	Calvo; 18604 (MA)								
Sutera foetida	South Africa; Cape, Clanwilliam District	–	T.M. Salter; JC3188 (MA)			–	–	–	–	–	–
Sutera foetida	South Africa; Cape, Oudeberg	Grid Ref. 3320 Montagu	J.P.H. Acocks; 2780 (K)								
Teedia lucida	–; –	–	n.a.; 20538 (K)								
Verbascum virgatum	Spain; Badajoz, Jerez de los Caballeros	–	Bote; F612 (K)								
Zaluzianskya violacea	South Africa; WesternCape Province, Beaufort West	-32.3604667 22.4872500	Medina; n.a. (MA)								–
Plantago chimera	–; –	–	Medina; LM6112 (MA)	AB223157.1	EU580431.1	KT177134.1	–	KX037415.1	KJ579147.1	EU036242.1	–

Table S2. Primers and specific protocols used for PCR amplification of markers.

	Primer Sequence				Amplifying Protocol					
	Region	Primer Name	Sequence (5'-3')	Reference	Prior denature	Denature	Annealing	Extension	Final Extension	Cycles
ITS	ITS	ITS-u1-fwd	GGA AGG AGA AGT CGT AAC AAG G	Cheng et al. 2015	95ºC, 5min	95ºC, 60s	52-55ºC, 90s	72ºC, 45s	72ºC, 10m	30-40
		ITS4-rv	TCC TCC GCT TAT TGA TAT GC	White et al. 1990						
	ITS plant specific (p5-4) and (p5-2, p3-4)	ITS-p2-rv	GCC RAG ATA TCC GTT GCC GAG	Cheng et al. 2015	95ºC, 5min	95ºC, 60s	52-55ºC, 90s	72ºC, 45s	72ºC, 10m	30-40
		ITS-p3-fwd	YGA CTC TCG GCA ACG GAT A	Cheng et al. 2015						
		ITS-p4-rv	CCG CTT AKT GAT ATG CTT AAA	Cheng et al. 2015						
		ITS-p5-fwd	CCT TAT CAY TTA GAG GAA GGA G	Cheng et al. 2015						
Chloroplast	*ndhF–rpl32*	rpL32-R	CCA ATA TCC CTT YYT TTT CCA A	Shaw et al. 2007	95ºC, 5min	95ºC, 60s	52ºC, 90s	72ºC, 45s	72ºC, 10m	35
		ndhF	GAA AGG TAT KAT CCA YGM ATA TT	Shaw et al. 2007	95ºC, 5min	95ºC, 60s	53ºC, 60s	65ºC, 120s	65ºC, 10m	40
	$trnL^{UAA}$–$trnF^{GAA}$	c B49317-fw	CGA AAT CGG TAG ACG CTA CG	Taberlet et al. 1991	95ºC, 5min	95ºC, 60s	52ºC, 90s	72ºC, 45s	72ºC, 10m	35-45
		f A50272-rv	ATT TGA ACT GGT GAC ACG AG	Taberlet et al. 1991	94ºC, 2min	94ºC, 30s	52ºC, 60s	72ºC, 60s	72ºC, 10m	30
	$trnT^{UGU}$–$trnL^{UAA}$	trnA2-fw	CAA ATG CGA TGC TCT AAC CT	Cronn et al., 2002	95ºC, 5min	95ºC, 60s	52ºC, 90s	72ºC, 45s	72ºC, 10m	35
		b A49291-rv	TCT ACC GAT TTC GCC ATA TC	Taberlet et al. 1991	94ºC, 1min	95ºC, 60s	54ºC, 120s	72ºC, 120s	72ºC, 10m	30-40
	$trnS^{GCU}$– $trnG^{GCC}$	$trnS^{GCU}$-fwd	GCC GCT TTA GTC CAC TCA GC'	Hamilton 1999	95ºC, 5min	95ºC, 60s	50-55ºC, 90s	72ºC, 45s	72ºC, 10m	30-40
		$trnG^{GCC}$-rv	GAA CGA ATC ACA CTT TTA CCA C	Hamilton 1999						
	rpl32-$trnL^{UAG}$	$trnL^{UAG}$-fwd	CTG CTT CCT AAG AGC AGC GT	Shaw et al. 2007	95ºC, 5min	95ºC, 60s	52ºC, 90s	72ºC, 45s	72ºC, 10m	35
		rpl32-F-rv	CAG TTC CAA AAA AAC GTA CTT C	Shaw et al. 2007						
	rps16 intron	rpS16F	AAA CGA TGT GGT ARA AAG CAA C	Shaw et al. 2005	95ºC, 5min	95ºC, 60s	52ºC, 90s	72ºC, 45s	72ºC, 10m	35
		rpS16R	AAC ATC WAT TGC AAS GAT TCG ATA	Shaw et al. 2005						
	psbJ–petA	psbJ-fwd	ATA GGT ACT GTA RCY GGT ATT	Shaw et al. 2007	95ºC, 5min	95ºC, 60s	51-53ºC, 90s	72ºC, 45s	72ºC, 10m	30-40
		petA-rv	AAC ART TYG ARA AGG TTC AAT T	Shaw et al. 2007						
	psbJ–petA Camptoloma specific	psbJ-petA_fwd	ACA CCC CTT TCT TGT GTC GAA GAC	This study.	95ºC, 5min	95ºC, 60s	51-53ºC, 90s	72ºC, 45s	72ºC, 10m	30-40
		psbJ-petA_rv	GAA CGG ACT ATA AAT GCG GGG AAC A	This study.						

	petB–petD	PlpetB1365F-fwd PlpetD738R-rv	TTG ACY CGT TTT TAT AGT TTA C AAT TTA GCY CTT AAT ACA GG	Löhne et al. 2005 Löhne et al. 2005	96ºC, 1min	95ºC, 60s	49ºC, 120s	72ºC, 120s	72ºC, 10m	40
	petB–petD *Camptoloma specific*	petBD_fwd petBD_rv	TCT GCC ACA TTG GAA TTC ACA ACC A AGG GGT TAC ACG GTG GTC GG	This study. This study.	96ºC, 1min	95ºC, 60s	49ºC, 120s	72ºC, 120s	72ºC, 10m	40

Table S3. Summary statistics of the genetic regions studied for the "outgroup" dataset and each individual species of *Camptoloma* ("population-level" datasets); see text for more details.

"Outgroup" dataset (outgroups + 2 specimens of each Camptoloma species)	**ITS**	***rpl32-ndhF***	***petB-petD***	***psbJ-petA***	***rps16 Intron***	***trnL-trnF***	***trnS-trnG***	***trnT-trnL***
Nº of sequences	42	44	35	37	41	38	39	41
Nº of taxa	24	26	21	21	24	22	23	22
Length matrix alignment (Nº of base pairs)	721	785	1058	1256	961	1054	1055	1011
Variable sites (Nº and %)	4, 0.55	259, 33.00	383, 36.20	484, 35.54	307, 31.95	281, 26.66	419, 39.72	333, 32.94
Nº of parsimony-informative sites	231	169	159	239	184	170	216	191

Camptoloma canariense	**ITS**	***rpl32-ndhF***	***petB-petD***	***psbJ-petA***	***rps16 Intron***	***trnL-trnF***	***trnS-trnG***	***trnT-trnL***
Nº of sequences	8	8	8	8	8	8	8	8
Nº of populations	4	4	4	4	4	4	4	4
Length matrix alignment (Nº of base pairs)	642	623	932	1063	822	926	731	676

Variable sites (N° and %)	0, 0	5, 0.80	0, 0	2, 0.19	3, 0.36	0, 0	2, 0.27	1, 0.15
N° of parsimony-informative sites	0	5	0	2	3	0	1	1

Camptoloma lyperiiflorum	**ITS**	***rpl32-ndhF***	***petB-petD***	***psbJ-petA***	***rps16 Intron***	***trnL-trnF***	***trnS-trnG***	***trnT-trnL***
N° of sequences	17	17	15	13	17	13	14	15
N° of populations	6	6	6	6	6	6	6	6
Length matrix alignment (N° of base pairs)	650	679	932	1093	824	932	754	691
Variable sites (N° and %)	14, 2.15	30, 4.42	15, 1.61	28, 2.56	7, 0.85	5, 0.54	11. 1.46	6, 0.87
Parsimony-Informative Characters	15	5	6	6	3	1	4	1

Camptoloma rotundifolium	**ITS**	***rpl32-ndhF***	***petB-petD***	***psbJ-petA***	***rps16 Intron***	***trnL-trnF***	***trnS-trnG***	***trnT-trnL***
N° of sequences	10	10	7	10	7	7	7	8
N° of populations	5	5	5	5	5	5	5	5
Length matrix alignment (N° of base pairs)	649	615	932	1106	824	927	732	673
Variable sites (N° and %)	4, 0.62	2, 0.33	0, 0	24, 2.22	3, 0.36	1. 0.12	5, 0.68	0, 0
N° of parsimony-informative sites	1	1	0	5	0	0	1	0

Table S4. Geographic distribution of all genera of Scrophulariaceae included in our study. Outgroup genera were represented by one or two species, whose distribution may not cover the entire distribution of the genus (e.g. *Buddleja*, *Myoporum*).

	Genus	Species used	Represented distribution in our DNA samples	Genus Distribution
Ingroup				
Africa	*Camptoloma*	*canariense* *rotundifolium* *lyperiiflorum*	Gran Canaria Namibia Yemen, Oman	
Outgroups				
South Africa	*Aptosimum*	*arenarium*	Namibia	
	Jamesbrittenia	*heucherifolia*	Namibia	
	Manulea	*crassifolia* *pusilla* *glandulosa*	South Africa South Africa, South Namibia West half of South Africa	
	Dermatobotrys	*saundersii*	South Africa	
	Freylinia	*undulata*	South Africa	
	Glumicalyx	*lesuticus*	East half of South Africa	

	Gomphostigma	*virgatum*	South Africa, Namibia	
	Lyperia	*antirrhinoides*	West half of South Africa	
	Peliostomum	*virgatum cf.*	West half of South Africa	
	Phygelius	*capensis*	East half of South Africa	
	Oftia	*africana*	West half of South Africa	
	Pseudoselago	*sp.*	West half of South Africa	
	Teedia	*lucida*	South Africa	
	Sutera	*foetida*	West half of South Africa	
	Zaluzianskya	*violacea*	West side of South Africa	
Africa	*Hebenstretia*	*My sequence/ Genbank chimera mix*	South Africa, East Africa	
The Americas	*Emorya*	*suaveolens*	Mexico	

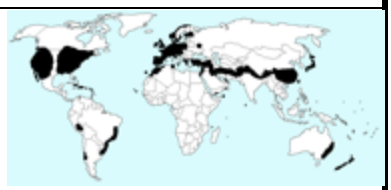
	Leucophyllum	*texanum* (synonym *frutescens)*	South USA, Mexico	
	Capraria	*biflora*	South USA, Central America, South America	
World Wide	*Buddleja*	*colvilei* *davidii* *megalocephala* *salvifolia* *tucumanensis*	Asia Asia Central America South East Africa South America	
	Limosella	*aquatica*	Europe, North America, Asia	
	Plantago	*Genbank chimera*	Unknown	
	Scrophularia	*arguta*	South coastline of Iberian peninsula, Canaria Islands, north African coastline from west to east, Yemen/Oman coastline	
	Verbascum	*virgatum*	Europe	
Australia and Pacific Islands	*Myoporum*	*sandwicense* *tetrandrum*	Pacific Islands (Hawaii)	

Table S5: Climatic variables from WorldClim2 (Fick and Hijmans 2017) with acronyms and units used in the study

VARIABLE NAME	SHORT HAND NAME AND UNITS
Minimum temperature for each month	tmin (ºc)
Maximum temperature for each month	tmax (ºc)
Mean temperature for each month	tmean (ºc)
Precipitation for each month	prec (mm)
Solar radiation for each month	srad (kJ m-2 day-1)
Wind speed for each month	wind (m/s)
Water vapour pressure for each month	vapr (kPa)
Annual Mean Temperature	BIO1 (ºc)
Mean Diurnal Range (Mean of monthly (max temp - min temp))	BIO2 (ºc)
Isothermality (BIO2/BIO7) × 100	BIO3 (ºc) × 100
Temperature Seasonality (standard deviation ×100)	BIO4 (ºc)
Max Temperature of Warmest Month	BIO5 (ºc)
Min Temperature of Coldest Month	BIO6 (ºc)
Temperature Annual Range (BIO5-BIO6)	BIO7 (ºc)
Mean Temperature of Wettest Quarter	BIO8 (ºc)
Mean Temperature of Driest Quarter	BIO9 (ºc)
Mean Temperature of Warmest Quarter	BIO10 (ºc)
Mean Temperature of Coldest Quarter	BIO11 (ºc)
Annual Precipitation	BIO12 (mm)
Precipitation of Wettest Month	BIO13 (mm)
Precipitation of Driest Month	BIO14 (mm)
Precipitation Seasonality (Coefficient of Variation)	BIO15 (mm)
Precipitation of Wettest Quarter	BIO16 (mm)
Precipitation of Driest Quarter	BIO17 (mm)

Precipitation of Warmest Quarter	BIO18 (mm)
Precipitation of Coldest Quarter	BIO19 (mm)

Table S6. Factor scores of the climatic variables in the PCA analysis using the full habitat suitability model for each present-day month. Only variables with a factor score > 0.6 with the first two PCA axes were included in the final model. Variables that had similar factor scores were clustered together, with the highest-scoring variable chosen to represent the cluster in the model. ***** marks the monthly climatic variables included in the "full dataset" model.

Time Period		Variables	PC1	PC2
Month	January	wind*	0.348	-0.698
		vapr*	-0.818	0.425
		tmin	-0.962	-0.158
		tmax	-0.964	-0.200
		tmean*	-0.976	0.181
		srad*	-0.823	-0.422
		prec*	-0.355	0.784
	February	wind*	0.455	-0.669
		vapr*	-0.774	0.501
		tmin	-0.962	-0.165
		tmax	-0.958	-0.223
		tmean*	-0.973	-0.196
		srad*	-0.782	-0.507
		prec*	-0.320	0.838
	March	wind*	0.414	0.760
		vapr*	-0.678	-0.633
		tmin	-0.969	0.116
		tmax	-0.953	0.228
		tmean*	-0.978	0.275
		srad*	-0.643	0.689
		prec*	-0.223	-0.891
	April	wind*	-0.006	0.855
		vapr*	-0.271	-0.886
		tmin	-0.950	-0.237
		tmax	-0.974	-0.029
		tmean*	-0.987	-0.136

		srad*	-0.639	0.678
		prec*	0.216	-0.875
	May	wind	0.633	0.570
		vapr*	-0.233	-0.889
		tmin	0.775	-0.612
		tmax*	0.897	-0.359
		tmean	0.861	-0.498
		srad*	0.789	0.392
		prec	-0.534	-0.662
	June	wind	0.725	0.400
		vapr	-0.183	-0.893
		tmin	0.847	-0.508
		tmax*	0.934	-0.231
		tmean	0.914	-0.376
		srad*	0.794	0.299
		prec*	-0.403	-0.744
	July	wind	0.735	0.351
		vapr*	-0.115	-0.897
		tmin	0.916	-0.329
		tmax	0.902	-0.323
		tmean*	0.931	-0.334
		srad*	0.794	0.310
		prec	-0.376	-0.785
	August	wind	0.695	0.394
		vapr	-0.167	-0.886
		tmin	0.882	-0.393
		tmax	0.874	-0.397
		tmean*	0.905	-0.407
		srad*	0.798	0.375
		prec*	-0.464	-0.737
	September	wind	0.561	0.568
		vapr*	-0.088	-0.899
		tmin	0.850	-0.432

		tmax*	0.902	-0.360
		tmean	0.909	-0.404
		srad*	0.770	0.390
		prec*	-0.470	-0.767
	October	wind*	0.0878	0.806
		vapr*	-0.518	-0.748
		tmin	-0.955	-0.049
		tmax	-0.970	0.141
		tmean*	-0.987	0.061
		srad*	-0.742	0.411
		prec*	0.003	-0.907
	November	wind*	0.279	-0.685
		vapr*	-0.745	0.544
		tmin	-0.975	-0.031
		tmax	-0.949	-0.259
		tmean*	-0.978	-0.159
		srad*	-0.800	-0.424
		prec*	-0.205	0.853
	December	wind*	0.332	-0.648
		vapr*	-0.812	0.447
		tmin	-0.979	-0.069
		tmax	-0.952	-0.257
		tmean*	-0.977	-0.172
		srad*	-0.822	-0.416
		prec*	-0.304	0.804

Table S7. Factor scores of the climatic variables in the environmental PCA analysis (ePCA)

Variables	PC1	PC2
BIO1	0.25	-0.04
BIO2	-0.07	-0.37
BIO3	0.09	-0.33
BIO4	-0.21	0.06
BIO5	0.20	-0.19
BIO6	0.24	0.10
BIO7	-0.16	-0.27
BIO8	0.22	-0.16
BIO9	0.21	0.20
BIO10	0.24	0.01
BIO11	0.25	-0.03
BIO12	-0.23	0.20
BIO13	-0.22	0.08
BIO14	0.07	0.24
BIO15	-0.14	-0.29
BIO16	-0.24	0.06
BIO17	0.10	0.26
BIO18	-0.08	-0.20
BIO19	-0.19	0.21
tmin yearly	0.24	0.11
tmax yearly	0.21	-0.20
tmean yearly	0.25	-0.04
srad yearly	0.19	0.06
prec yearly	-0.23	0.16
wind yearly	0.07	0.37
vapr yearly	0.23	0.14

Code

Code from this study can be accessed via email: vickycul@hotmail.com

ARTICLES PUBLISHED OUTWITH THE Ph.D.

frontiers in Genetics

ORIGINAL RESEARCH
published: 01 May 2015
doi: 10.3389/fgene.2015.00154

Living on the edge: timing of Rand Flora disjunctions congruent with ongoing aridification in Africa

Lisa Pokorny[1], Ricarda Riina[1], Mario Mairal[1], Andrea S. Meseguer[2], Victoria Culshaw[1], Jon Cendoya[1], Miguel Serrano[3], Rodrigo Carbajal[3], Santiago Ortiz[3], Myriam Heuertz[4, 5, 6] and Isabel Sanmartín[1]**

[1] Real Jardín Botánico (RJB-CSIC), Madrid, Spain, [2] INRA, UMR 1062, Centre de Biologie pour la Gestion des Populations (INRA, IRD, CIRAD, Montpellier SupAgro), Montferrier-sur-Lez, France, [3] Department of Botany, Pharmacy School, University of Santiago de Compostela, Santiago de Compostela, Spain, [4] Forest Research Centre (INIA-CIFOR), Madrid, Spain, [5] INRA, BIOGECO, UMR 1202, Cestas, France, [6] University of Bordeaux, BIOGECO, UMR 1202, Talence, France

The Rand Flora is a well-known floristic pattern in which unrelated plant lineages show similar disjunct distributions in the continental margins of Africa and adjacent islands—Macaronesia-northwest Africa, Horn of Africa-Southern Arabia, Eastern Africa, and Southern Africa. These lineages are now separated by environmental barriers such as the arid regions of the Sahara and Kalahari Deserts or the tropical lowlands of Central Africa. Alternative explanations for the Rand Flora pattern range from vicariance and climate-driven extinction of a widespread pan-African flora to independent dispersal events and speciation *in situ*. To provide a temporal framework for this pattern, we used published data from nuclear and chloroplast DNA to estimate the age of disjunction of 17 lineages that span 12 families and nine orders of angiosperms. We further used these estimates to infer diversification rates for Rand Flora disjunct clades in relation to their higher-level encompassing lineages. Our results indicate that most disjunctions fall within the Miocene and Pliocene periods, coinciding with the onset of a major aridification trend, still ongoing, in Africa. Age of disjunctions seemed to be related to the climatic affinities of each Rand Flora lineage, with sub-humid taxa dated earlier (e.g., *Sideroxylon*) and those with more xeric affinities (e.g., *Campylanthus*) diverging later. We did not find support for significant decreases in diversification rates in most groups, with the exception of older subtropical lineages (e.g., *Sideroxylon*, *Hypericum*, or *Canarina*), but some lineages (e.g., *Cicer*, *Campylanthus*) showed a long temporal gap between stem and crown ages, suggestive of extinction. In all, the Rand Flora pattern seems to fit the definition of biogeographic pseudocongruence, with the pattern arising at different times in response to the increasing aridity of the African continent, with interspersed periods of humidity allowing range expansions.

Keywords: Africa, historical biogeography, climate change, diversification rates, long-distance dispersal, Rand Flora, vicariance

OPEN ACCESS

Edited by:
James Edward Richardson, Royal Botanic Garden Edinburgh, UK

Reviewed by:
Thomas L. P. Couvreur, Institut de Recherche pour le Développement, Cameroon
Lars Chatrou, Wageningen University, Netherlands

****Correspondence:***
Lisa Pokorny and Isabel Sanmartín, Real Jardín Botánico (RJB-CSIC), Plaza de Murillo 2, 28014 Madrid, Spain
pokorny@rjb.csic.es; isanmartin@rjb.csic.es

Specialty section:
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics

Received: *06 October 2014*
Accepted: *05 April 2015*
Published: *01 May 2015*

Citation:
Pokorny L, Riina R, Mairal M, Meseguer AS, Culshaw V, Cendoya J, Serrano M, Carbajal R, Ortiz S, Heuertz M and Sanmartín I (2015) Living on the edge: timing of Rand Flora disjunctions congruent with ongoing aridification in Africa. Front. Genet. 6:154. doi: 10.3389/fgene.2015.00154

Introduction

Large-scale biodiversity patterns have intrigued naturalists since the eighteenth century (Forster, 1778; von Humboldt and Bonpland, 1805; Wallace, 1878; Fischer, 1960; Stevens, 1989;

Lomolino et al., 2010). Recognizing that spatial variation in environmental variables such as temperature or precipitation is insufficient to explain such patterns, more integrative explanations that emphasize the role of both environmental and evolutionary factors have recently been advanced (Qian and Ricklefs, 2000; Wiens and Donoghue, 2004; Jablonski et al., 2006). As Wiens and Donoghue (2004) state "environmental variables cannot by themselves increase or decrease local or regional species richness"; only evolutionary processes such as dispersal, speciation and extinction can. Therefore, reconstructing rates of dispersal, speciation, and extinction across the component lineages of a biota might help us understand how assembly took place across space and through time (Pennington et al., 2004; Ricklefs, 2007; Wiens, 2011). Moreover, understanding patterns of biotic assembly is a pressing goal in biodiversity research at a time when nearly one tenth of species on Earth are projected to disappear in the next hundred years (Maclean and Wilson, 2011).

Africa is a continent especially interesting to study patterns of biotic assembly. On one hand, African tropical regions are comparatively species-poorer than regions situated in the same equatorial latitudes in the Neotropics and Southeast Asia (Lavin et al., 2001; Couvreur, 2015), which has led to the continent being referred to as the "odd man out" (Richards, 1973). On the other, Africa offers some extraordinary examples of continent-wide disjunctions. For example, tropical rainforests in Africa appear in two main blocks, the West-Central Guineo-Congolian region and the coastal and montane regions of East Africa, now separated by a 1000 Km-wide arid corridor (Couvreur et al., 2008). Another prime example is the so called *Rand Flora* (RF), a biogeographic pattern in which unrelated plant lineages show comparable disjunct distributions with sister taxa occurring on now distantly located regions in the continental margins of Africa: Macaronesia-northwest Africa, Western African mountains, Horn of Africa-South Arabia (including the Island of Socotra), Eastern Africa (incl. Madagascar), and Southern Africa (Christ, 1892; Lebrun, 1947, 1961; Quézel, 1978; Andrus et al., 2004; Sanmartín et al., 2010; **Figure 1**). All RF lineages share sub-humid to xerophilic affinities, so that the tropical lowlands of Central Africa and the large Sahara and Arabian deserts in the north or the Namib and Kalahari deserts in the south presumably constitute effective climatic barriers to their dispersal.

Swiss botanist K. H. H. Christ (1892) first referred to "cette flore *marginale* de l'Afrique," that is "this *marginal* African flora," in a note addressing the role the so called *ancient African flora* played on European floras, with emphasis on the Mediterranean biome. Later, in his "Die Geographie der Farne" (i.e., "The Geography of Ferns"; Christ, 1910), he very aptly named this geographic pattern "Randflora" (see pp. 259–275), where the Germanic word "Rand" stands for rim, edge, border, margin (see **Figure 1** inset), noting its similarities with Engler's "afrikanisch-makaronesische Element" (Engler, 1879, 1910; see pp. 76 in the former and pp. 983–984 and 1010 in the latter), that is, an "Afro-Macaronesian element" linking disjunct xerophilic taxa found in the continental margins of Africa and its adjacent islands (e.g., Canary Islands, Cape Verde, etc.).

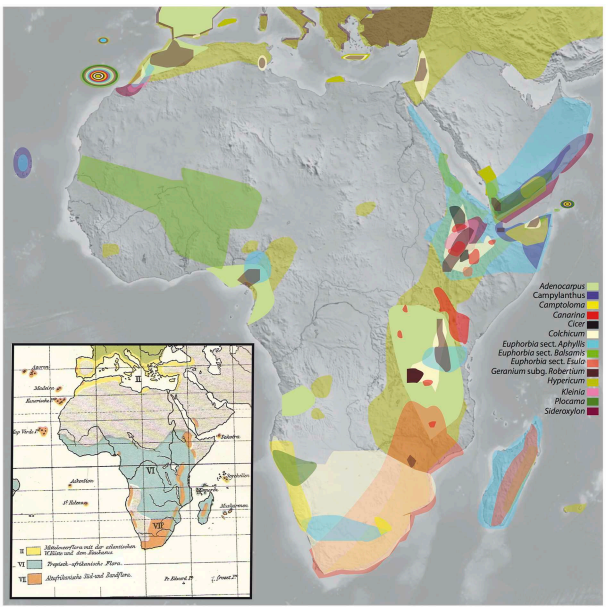


FIGURE 1 | Rand Flora disjunction pattern as evidenced by angiosperm plant lineages analyzed for this study. The inset shows K.H.H. Christ's (1910) depiction of "cette flore *marginale* de l'Afrique" or "Randflora" (in orange color), note their similar geographic limits. Taxa: *Adenocarpus* (Fabaceae), *Camptoloma* (Scrophulariaceae), *Campylanthus* (Plantaginaceae), *Canarina* (Platycodoneae, Campanulaceae), *Cicer* (Fabaceae), *Colchicum* (Colchicaceae), *Euphorbia* subgen. *Athymalus* (sects. *Anthacanthae* and *Balsamis*; Euphorbiaceae), *Euphorbia* subgen. *Esula* (sect. *Aphyllis*), *Euphorbia* subgen. *Esula* (African clade of sect. Esula), *Geranium* subgen. *Robertium* (Geraniaceae), *Hypericum* (Hypericaceae), *Kleinia* (Asteraceae), *Plocama* (Rubiaceae), and *Sideroxylon* (Sideroxyleae, Sapotaceae).

Historical explanations for this pattern and, in particular, its temporal framework, its exact boundaries, and the ecology of the plants involved have varied through these past two centuries. The early view (Engler, 1879, 1910; Christ, 1892, 1910) was one of a pan-African flora found throughout the continent that became restricted to its margins as a result of major climate changes (i.e., increasing aridification) throughout the *Tertiary* (i.e., the Cenozoic Period, 66.0–2.58 Ma). Lebrun (1947; see pp. 134–137), and later Monod (1971, p. 377) and Quézel (1978, p. 511), interpreted Christ's *ancient African flora* as a complex ensemble that had experienced alternating expansions and contractions through time, having had a chance to spread across northern Africa during favorable moments in the Miocene and needing to retract at the end of the Neogene (i.e., Pliocene): a further increase in aridity at the beginning of Pleistocene glaciations would have confined relictual or vicariant taxa to Macaronesia, northwest Africa and Arabia. Axelrod and Raven (1978) explained some of these disjunctions in relation to a more ancient, widespread Paleogene flora of subtropical origin that covered the entire African continent at the beginning of the Cenozoic, and that was decimated by successive events of aridification, of which the relict floras of Macaronesia, the Cape Region, and the Afromontane forests in eastern and western Africa would be remnants. Bramwell (1985) explains this pattern in terms of

pan-biogeographic "general tracks" that connect what would be the remains of an ancient flora that extended across the Mediterranean and Northern Africa in the Miocene, and whose vestiges could be found in the Macaronesian laurisilva and a few enclaves in the island of Socotra, the Ethiopian Highlands and southern Yemen.

These authors share a vicariant perspective and presume RF lineages were part of a widespread pan-African *Tertiary* flora that became fragmented by the appearance of climatic barriers (i.e., aridification), leaving relictual lineages with reduced distributions at "refugia" in the margins of Africa (i.e., "continental" islands). This "refugium" idea rests on the fact that many of these RF regions—Macaronesia, the South African Cape region, and the semi-arid regions of Eastern Africa and Southern Arabia (e.g., Ethiopia, Yemen, Socotra)—harbor a large number of endemic species, when compared to neighboring areas. Moreover, the "fragmentation-refugium" hypothesis implies the disappearance, possibly by extinction, of RF lineages from part of their distributional range (e.g., across the Sahara in central Northern Africa), which is consonant with the "climatic vicariance" concept (Wiens, 2004): an environmental change creates conditions within a species' geographic range that are outside the ancestral climatic tolerances; individuals are unable to persist and the species' geographic range becomes fragmented.

The alternative explanation is one of independent dispersal (immigration) events among geographically isolated regions and subsequent speciation *in situ*. In this framework, divergence events need not be congruent across lineages, since long-distance dispersal (LDD) events are highly stochastic in nature (Nathan, 2006). Asides from transoceanic dispersal—which has been postulated in the case of *Aeonium* (Kim et al., 2008), *Geranium* (Fiz et al., 2008), and other RF lineages (Andrus et al., 2004) based on molecular phylogenetic evidence—, cross-continent LDD dispersal is also possible: published examples favoring cross-continent LDD include *Senecio*, with a disjunct distribution between Macaronesia-Northern Africa and South Africa (Coleman et al., 2003; Pelser et al., 2012). Moreover, dispersal does not necessarily imply long-distance migration events. In some cases, dispersal across intermediate areas that act as "stepping stones" or "land bridges" could have been possible. For example, the presence of isolated mountain ranges (offering suitable habitats) throughout the Sahara, such as the Tibesti and Hoggar massifs, could have allowed this short or medium-range dispersal in *Campanula* (Alarcón et al., pers. comm.). Correspondingly, some RF lineages might have used the Arabian Plate as a land bridge to reach East Africa (*Campanula*, Roquet et al., 2009; *Hypericum*, Meseguer et al., 2013), and others may have benefited from the new habitats offered by the Pliocene uplift of the Eastern Arc Mountains to migrate to or from South Africa (Meseguer et al., 2013).

Discriminating between climate-driven vicariance vs. independent dispersal events between geographically isolated regions requires framing the evolution of disjunct lineages on a temporal scale (Sanmartín, 2014). On the other hand, to unravel the origin of a biota or biome, a meta-analysis across dated phylogenies of multiple non-nested clades is needed (Pennington et al., 2010; Wiens, 2011; Couvreur, 2015). Sanmartín et al. (2010) carried out a meta-analysis of 13 lineages to infer relative rates of historical dispersal among RF regions (Macaronesia, Eastern Africa-Southern Arabia, and Southern Africa) and found the highest rate of biotic exchange between east and west Northern Africa, across the Sahara. However, they did not integrate absolute estimates of lineage divergences in their inference, since very few RF lineages (e.g., Roquet et al., 2009) had been dated at the time.

In this study, we estimate time divergences for up to 13 plant lineages (**Table 1**) displaying RF disjunct distributions (**Figure 1**), and use published divergence times for four other lineages (see Materials and Methods), in order to provide a much-needed temporal framework for this pattern. An extensive description of each of these lineages, geographic distributions and phylogenetic relationships is provided in Supplementary Materials. We also frame these disjunctions in the context of major climatic and geological events in the history of Africa (see summary below) and estimate net diversification rates in an attempt to address the role that evolutionary processes, such as climate-driven extinction, may have played in the formation of the African RF pattern.

Materials and Methods

Study Area: African Climate through Time

To understand biogeographic patterns in the African flora, it is necessary to briefly review the climatic and geological history that might have influenced the evolution of African plant lineages. Extensive reviews of African climatic and vegetation history can be found in Axelrod and Raven (1978); van Zinderen Bakker (1978); Maley (1996, 2000); Morley (2000); Jacobs et al. (2010), Plana (2004), and Bonnefille (2011), among others.

During the Late Mesozoic, Africa was part of the supercontinent Gondwana, located in the southern hemisphere, and enjoyed a relatively humid and temperate climate (Raven and Axelrod, 1974). After breaking up from South America ca. 95 Ma, Africa started moving northwards toward the equatorial zone (**Figure 2A**). The result was a general trend toward continental aridification in which different regions became arid or wet at alternative times (**Figure 2B**, Senut et al., 2009). Paleocene Africa (66–56 Ma) was mainly wet and warm, characterized by a major diversification in the West African flora (Plana, 2004). A global increase in temperatures in the Eocene (56–33.9 Ma) led to increased aridity in Central Africa, with a rainforest-savannah mosaic in the Congo region. This was followed by a global cooling event at the Eocene-Oligocene boundary (33.9 Ma), which led again to aridification and major extinction but did not change biome composition (Axelrod and Raven, 1978).

The Early Miocene (23–16 Ma) was warm and humid, with wide extension of rainforests, from the northern Sahara to parts of Southern Africa. The Mid Miocene (16–11.6 Ma) was a period of major changes in climate and topography. A combination of factors, including the gradual uplift of Eastern Africa, the successive closure of the Tethys seaway in the north, and the expansion of the East Antarctic ice sheet in the south (Trauth et al., 2009), led to a general intensification of the aridification process, though it was not homogeneous across the continent. Geological and paleontological evidence suggest that now arid regions (e.g., northern Africa, Horn of Africa, Namib Desert) were during

TABLE 1 | Rand Flora disjunctions, encompassing (higher level) lineages, recent molecular phylogenetic studies, and molecular markers used in here.

Order	Family	Tribe (or else)	Genus	Subgenus	Section (or else)	Disjunction name	Dataset reference	Molecular marker	
								Nuclear	Chloroplast
Fabales	Fabaceae	Genisteae	*Adenocarpus*			*Ad. manii*	Cubas et al., 2010	ETS, ITS	*trn*LF
Saxifragales	Crassulaceae	*Aeonium* alliance	*Aeonium*			*Ae. leucoblepharum*	Mort et al., 2002, 2007	ITS	–
Malpighiales	Euphorbiaceae		*Euphorbia*	*Athymalus*	*Anthacanthae* *Balsamis*	*Eu. omariana* *Eu. balsamifera*	Peirson et al., 2013	ITS	*ndh*F
Malpighiales	Euphorbiaceae			*Esula*	*Aphyllis Esula*	*Eu. tuckeyana* *Eu. usambarica* *Eu. schimperiana*	Barres et al., 2011; Riina et al., 2013	ITS	*ndh*F
Asterales	Campanulaceae		*Campanula*		*Azorina* (clade)	*Ca. jacobaea*	Alarcón et al., 2013	–	*trn*LF, *pet*BD, *rpl*32–*trn*L, *trn*SG
Lamiales	Scrophulariaceae	Buddlejoideae (subfamily)	*Camptoloma*			*Cm. canariense* *Cm. rotundifolium*	Kornhall et al., 2001; Oxelman et al., 2005	–	*trn*LF, *ndh*F, *rps*16
Lamiales	Plantaginaceae	Globularieae	*Campylanthus*			*Cy. salsoloides*	Thiv et al., 2010	ITS	*atp*B-*rbc*L
Asterales	Campanulaceae	Platycodoneae	*Canarina*			*Cn. canariensis*	Mairal et al., 2015	ITS	*pet*BD, *psb*J, *trn*LF, *trn*SG
Fabales	Fabaceae	Vicioids (clade)	*Cicer*			*Ci. canariense*	Javadi et al., 2007	ETS, ITS	*trn*SG, *mat*K, *trn*AH, *trn*A-Leu
Liliales	Colchicaceae	Colchiceae	*Colchicum*			*Co. schimperianum*	Manning et al., 2007; del Hoyo et al., 2009	–	*trn*LF, *atp*B-*rbc*L, *rps*16
Geraniales	Geraniaceae		*Geranium*	*Robertium*		*G. robertianum*	Fiz et al., 2008	ITS	–
Malpighiales	Hypericaceae	Hypericeae	*Hypericum*		*Androsaemum* *Campylosporus*	*H. scopulorum* *H. quartinianum*	Meseguer et al., 2013	–	*trn*LF, *trn*SG
Asterlaes	Asteraceae	Senecioneae	*Kleinia*			*K. neriifolia*	Pelser et al., 2007	ITS	*trn*LF
Gentianales	Rubiaceae	Putorieae	*Plocama*			*Pl. pendula* *Pl. crocyllis*	Backlund et al., 2007	–	*rps*16, *trn*TF, *atp*B-*rbc*L
Ericales	Sapotaceae	Sideroxyleae	*Sideroxylon*			*S. spinosus*	Smedmark et al., 2006; Smedmark and Anderberg, 2007	–	*ndh*F, *trn*H–*psb*A, *trn*CD

GenBank numbers can be found in the references listed under column "Dataset reference."

this period more humid than they are today, whereas other now humid regions (e.g., Congo Basin) were much drier (**Figure 2B**). Desertification started in the southwest (Namib Desert) around 17–16 Ma ago, and proceeded eastward and northward. In Southern Africa, tropical to subtropical vegetation was replaced by wooded savannah during the lower Mid-Miocene (Senut et al., 2009). In Northern Africa, the earliest evidence of aridity in the Sahara region is from the Late Miocene (11.6–5.3 Ma), ca. 7–6 Ma (Senut et al., 2009; **Figure 2B**). In Central Africa, a semi-arid desert ("Miocene Congo Desert," **Figure 2B**) occupied the region until the Mid Miocene, 13–12 Ma ago, when the Eastern African uplift and subsequent subsidence led to the establishment of the Congo River drainage and a general increase in humidity ("tropicalization"). Also in the Late Miocene, ca. 7–8 Ma, a new period of tectonic activity in Eastern Africa led to the uplift of the Eastern Arc Mountains and the uplands of West Central Africa (Cameroon volcanic line), which led to increasing aridity and the expansion of savannahs and grasslands in these regions (Sepulchre et al., 2006). Uplifting reached a maximum during the Plio-Pleistocene and led to the formation of the Ethiopian Highlands and the desertification of low-lying areas in the Horn of Africa (Senut et al., 2009). From the Late Pliocene to the Holocene, the alternation of glacial-and interglacial periods seems to have led to repeated contractions and expansions of distributional ranges across both subtropical and tropical taxa (Maley, 2000; Bonnefille, 2011). Some areas like the Saharan massifs of Tibesti and Hoggar or the Ennedi Mountains could have served as refuges during arid periods for subtropical taxa (Osborne et al., 2008), whereas the uplands of Upper and Lower Guinea and the east of the Congo Basin, the Albertine Rift, or the Eastern Arc Mountains could have played the same role for tropical plant taxa (Maley, 1996; **Figure 2B**).

Taxon Sampling

We retrieved sequences from GenBank from existing studies (**Table 1**) for the following 13 lineages exhibiting a distribution congruent with the RF pattern (Andrus et al., 2004; Sanmartín et al., 2010): *Adenocarpus*, *Aeonium*, *Camptoloma*, *Campylanthus*, *Cicer*, *Colchicum*, *Euphorbia* sects. *Antachanthae*, *Aphyllis*, *Balsamis*, and *Esula*, *Geranium*, *Kleinia*, and *Plocama* (**Figure 3**).

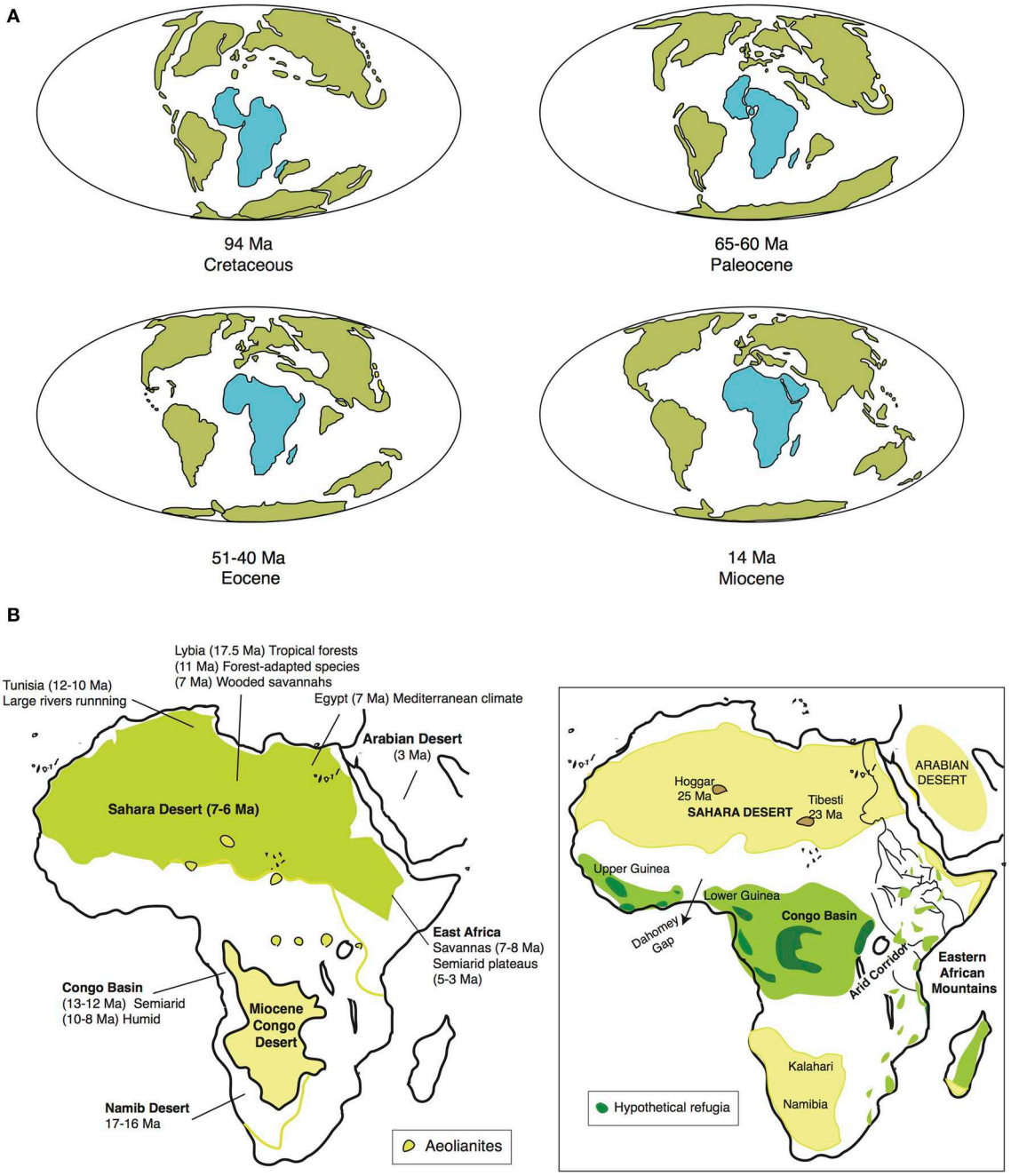


FIGURE 2 | (A) Tectonic fragmentation of the supercontinent Gondwana through time, showing Africa's drift northwards; and **(B)** main climatic events in Africa during Neogene (adapted from Senut et al., 2009): (B-left) Early Neogene Central Africa was more arid than North Africa, with a desert, semiarid region in the Congo Basin. Desertification started in southwest Africa in the Mid-Miocene, proceeding eastward and northward, and finalizing with the formation of the Sahara Desert. Conversely, Central Africa became tropical due to subsidence and Eastern African uplift. (B-right) Schematic representation of present-day vegetation belts, showing position of main deserts and rainforest refugia (Eastern Arc Mountains/Guineo-Congolian region (the latter fragmented into smaller refugia). Rand Flora lineages occupy the regions in the margin that are not deserts or rainforests, rarely some find refuge in mountain areas of North African Sahara (e.g., Tibesti and Hoggar Massifs).

We chose these lineages because sampling is nearly complete in most cases with very few to no missing taxa. Most of these RF taxa have been sequenced for several markers from the nuclear and chloroplast DNA regions. For each group we selected the markers with most sequences and tried representing both genomic compartments whenever possible. The sequences were aligned using the Opalescent package (Opal v2.1.0; Wheeler and Kececioglu, 2007) in Mesquite v3.01 (Maddison and Maddison, 2014) and

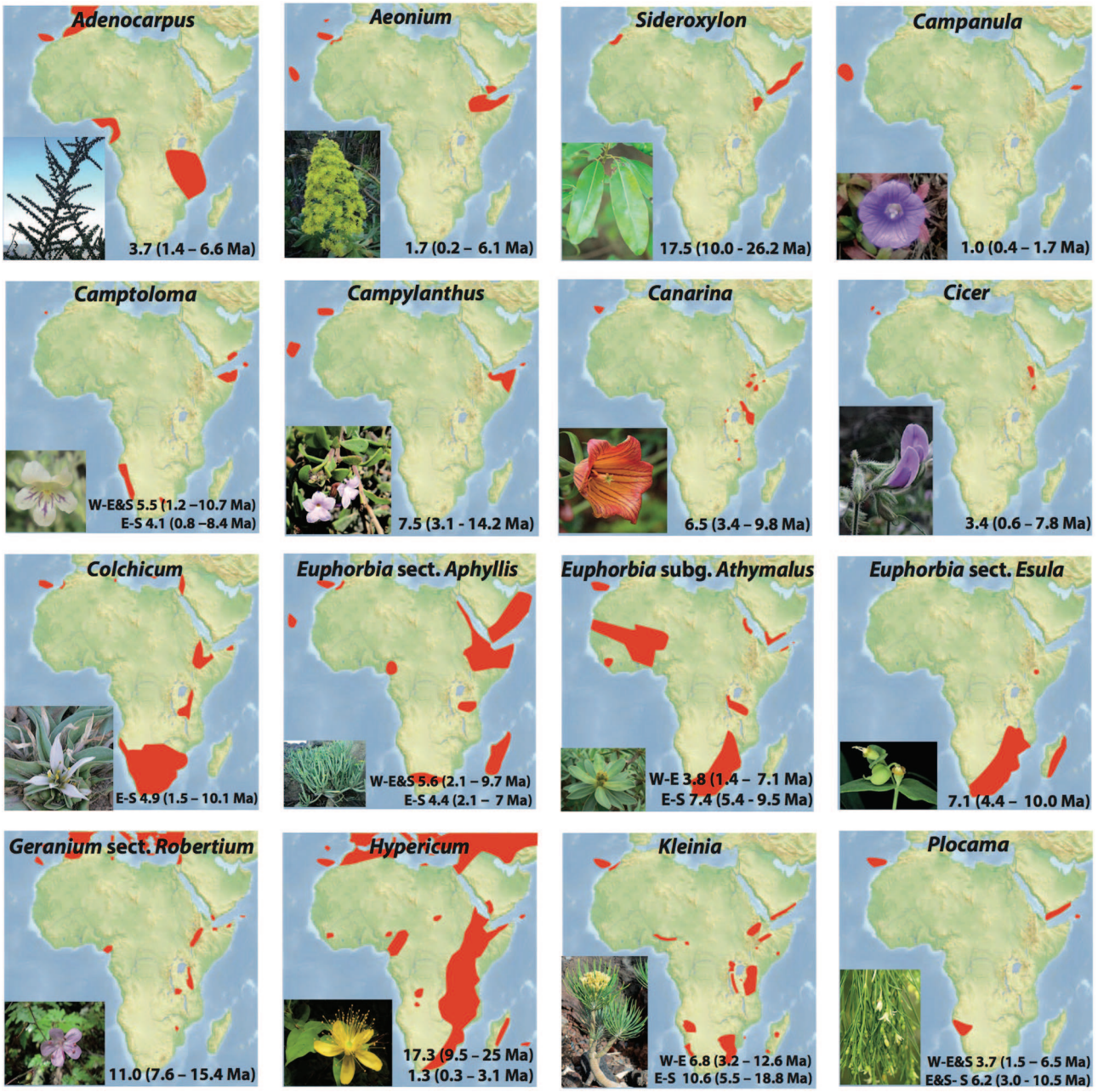


FIGURE 3 | Individual distributions and habit illustrations for 16 plant lineages exhibiting Rand Flora disjunctions. Estimated divergence times within each lineage correspond to the disjunctions represented in **Figures 4**, **5** and indicated in the MCC chronograms shown in **Figures S1–S16**. Taxa names correspond to those in **Table 1**.

manually adjusted in SE-AL v2.0a11 (Rambaut, 2002) using a similarity criterion, as recommended by Simmons (2004). For four other RF lineages —*Campanula* (Alarcón et al., 2013), *Canarina* (Mairal et al., 2015), *Hypericum* (Meseguer et al., 2013), and *Sideroxylon* (Stride et al., 2014)—we used recently published time estimates by our research team (except for *Sideroxylon*, which nonetheless used a dating approach similar to ours). Approximately 1600 sequences from ca. 675 taxa from 12 families and 9 orders of angiosperms were included in our study (**Table 1**).

Estimating Absolute Divergence Times

Divergence times were estimated under a Bayesian framework in BEAST v1.8 (Drummond et al., 2012). For each lineage, we constructed a dataset including the markers listed in **Table 1**, which were partitioned by genome (chloroplast vs. nuclear),

whenever possible. The best-fitting substitution model for each partition was selected using the Akaike Information Criterion implemented in MrModeltest v2.2 (Nylander, 2004) and run in PAUP* v4.0b (Swofford, 2002). The relaxed uncorrelated lognormal clock model (UCLD, Drummond et al., 2006) and a Yule speciation process as tree model were selected for all datasets based on preliminary explorations. MCMC searches were run 5×10^7 generations and sampled and logged every 2500th generation. We used Tracer v1.6 (Rambaut et al., 2013) to determine stationarity of the Markov chain and to verify that all parameters had large enough effective sampling sizes (ESS>200). TreeAnnotator v1.8.0 (Drummond et al., 2012) and FigTree v. 1.4.2 (Rambaut, 2009) were used respectively to generate and visualize the resulting maximum clade credibility (MCC) chronograms.

Calibration points for obtaining absolute divergence times were based on either the fossil record or on published secondary calibration constraints (**Table 2**). The latter were obtained from published dated phylogenies of datasets including our study groups (e.g., the family to which the genus belongs), and were assigned normal distribution priors (Ho and Phillips, 2009) in the BEAST analysis that encompassed the mean and the 95% highest posterior density (HPD) confidence interval (CI) from these studies [except in the case of time constrains from Bell et al. (2010), for which a lognormal distribution was used, since posterior estimates for a normal prior were not available]. For fossil calibration points we used a lognormal prior, since this distribution better represents the stratigraphic uncertainty associated to the fossil record (Ho and Phillips, 2009). The offset of the lognormal distribution was set to the upper bound of the stratigraphic period where the fossil was found, and the standard deviation (SD) and mean were set so that the 95% CI encompassed the lower and upper bound of the period (e.g., for Late Eocene *Hypericum antiquum* a lognormal distribution offset at 33.9 Myr, with mean = 1.0 and SD = 0.7, was used to cover the length of the period where the fossil was found, that is 33.9–37.2 Ma). A summary of time constraints used for each dataset and their provenance can be found in **Table 2**.

Diversification Analyses

We used divergence times estimated above to calculate absolute diversification rates in the aforementioned lineages. There have been numerous developments in macroevolutionary birth-death models that allow a more accurate estimation of extinction and speciation rates from dated molecular phylogenies, including episodic time-variable models and trait-dependent diversification models (Stadler, 2013; Morlon, 2014; Rabosky et al., 2014).

TABLE 2 | Time constraints and prior probability distributions imposed on constrained nodes to estimate divergence times in RF lineages.

Taxon set	Node constrained	Time constraint (Myr)			Dating reference	Figure/Table/P.
		Distribution (offset)	Mean	SD		
Adenocarpus	ROOT: Genisteae	Normal	19.5	3.8	Lavin et al., 2005	Table 2, node 32
Aeonium alliance	ROOT: *Aeonium* alliance	Normal	18.83	1.0	Kim et al., 2008	Figure 2C
E. subg. *Athymalus*	*Athymalus* w/o *E. antso*	Normal	10.78	2.0	Horn et al., 2014	Figure 2
sect. *Anthacanthae*	CROWN: *Athymalus*	Normal	24.56	5.0		Table 1
and sect. *Balsamis*	*Anthacanthae*	Normal	18.22	3.4		Table 1
	MRCA *Anthacanthae-Balsamis*	Normal	7.56	1.4		Figure 2
E. subg. *Esula*	MRCA *Aphyllis-Exiguae* II	Normal	10.36	2.3	Horn et al., 2014	Figure 2
sect. *Aphyllis*	CROWN: *Aphyllis*	Normal	7.37	2.0		Figure 2
E. subg. *Esula*	MRCA *Arvales-Esula*	Normal	10.98	2.4	Horn et al., 2014	Figure 2
sect. *Esula*	CROWN: *Esula*	Normal	8.6	2.4		Figure 2, node 5
(African clade)	*E. virgata* clade	Normal	5.4	1.4		Figure S2
Camptoloma	MRCA *Buddlejeae-Camptoloma*	Normal	20.0	6.0	Navarro-Pérez et al., 2013	Figure 2
	Buddlejeae	Normal	7.5	3.0		Figure 2
Campylanthus	MRCA *Digitalis-Plantago*	Lognormal (0.0)	38.0	0.2	Bell et al., 2010	Figure S11
	MRCA *Plantago-Aragoa**	Lognormal (7.1)	1.5	1.0	Thiv et al., 2010	P. 610
Cicer	CROWN: *Cicer*	Normal	14.8	5.0	Lavin et al., 2005	Figure 3, node 80
Colchicum	MRCA *Gloriosa-Colchicum*	Normal	43.3	7.0	Chacón and Renner, 2014	Figure 3, node 128/Table 2
Geranium subg. *Robertium*	MRCA *Pelargonium-Geranium*	Normal	28.0	3.0	Fiz et al., 2008	Figure 3, node D
	CROWN: *Robertium*§	Lognormal (7.25)	1.0	1.0		P. 329
Kleinia	ROOT: Asteraceae†	Lognormal (47.5)	10.0	0.75	Barres et al., 2013	P. 872
	Lordhowea insularis	Lognormal (0.0)	7.0	1.0	Pelser et al., 2010	Table 1
Plocama	MRCA *Putorieae-Paederieae*	Normal	34.4	5.5	Bremer and Eriksson, 2009	Table 1

*At least one node (preferably toward the root) was constrained in each phylogeny (**Figures S1–S16** show resulting chronograms explicitly stating any constrained nodes).*

** Plantaginacearumpollis miocenicus (Late Miocene, 10.3 Ma; Nagy, 1963; Doláková et al., 2011).*

§ Geranium cf. lucidum (Late Miocene, 7.246 Ma ± 0.005; Van Campo, 1989).

† Raiguenrayun cura (Middle Eocene, 47.5 Ma; Barreda et al., 2012).

However, these methods usually require both very large phylogenies (e.g., ≥100 tips) and a fairly complete sampling. We here chose a simpler approach, the "method-of-moments" estimator (Magallón and Sanderson, 2001), implemented in the R package GEIGER (Harmon et al., 2008). This method uses clade size (extant species number) and clade age (either crown or stem) to estimate net diversification rates (r = speciation minus extinction), under different values of background extinction or turnover rate (ε = extinction/speciation = 0.0, 0.5, and 0.9). Net diversification rates (*bd.ms* function in GEIGER) were here estimated for all RF disjunctions and for a series of successively encompassing clades (e.g., section, genus, tribe, subfamily, and so on) to detect possible rate shifts. Crown diversification rates could not be estimated for clades containing only two taxa because Magallón and Sanderson's formula ($r = [\log(n)–\log 2]/t$ in its simplest version, that is, with no extinction; for $\varepsilon > 0$ see formula number 7 in Magallón and Sanderson, 2001) results in zero in this case. In an attempt to counter this problem, clades containing two taxa were assigned a diversity value of 2.01, which permitted the estimation of net diversification rates (r).

Additionally, the probability of obtaining a clade with the same size and age as the RF disjunction, given the background diversification rate of the encompassing clade/s and at increasing extinction fractions ($\varepsilon = 0$, 0.5, and 0.9), was estimated with the *crown.p* function in GEIGER. We also estimated the 95% confidence interval of expected diversity through time (*crown.limits* function, GEIGER, $\varepsilon = 0$, 0.5, and 0.9) for a clade that diversifies with a rate equal to that of the family containing a RF disjunction with the highest diversification rate (i.e., Asteraceae); we then mapped RF lineages according to their crown or stem age and standing species diversity to assess which RF disjunct clades are significantly less diverse than expected given their stem and crown age in relation to the highest rate calculated for a RF family (Magallón and Sanderson, 2001; Warren and Hawkins, 2006).

Results

Divergence Times

Up to 21 disjunctions were identified and divergence times were estimated for 17 lineages exhibiting a geographic distribution consistent with the RF pattern (**Figures 3**, **4** and **Figures S1–S17**). These disjunctions represent two possible geographic splits: I) Eastern Africa (including the Eastern Arc Mountains, the Horn of Africa, and Southern Arabia) vs. Southern Africa (including southern Angola and Namibia and the Cape Flora region up to the Drakensberg Mountains), hereafter E-S, and II) Western Africa (including Macaronesia and NW Africa south to the Cameroon volcanic line) vs. Eastern Africa, (with or without S Africa), hereafter W-E(&S).

From youngest to oldest, E-S disjunctions (**Figure 4**) occur in *Plocama* (ca. 4 Ma between S African *Pl. crocyllis* on one side and, among other E African-S Arabian species, *Pl. yemenensis* and *Pl. tinctoria* on the other; **Figure 3** and **Figure S15**), *Camptoloma* (ca. 4 Ma between E African *Cm. lyperiiflorum* and S African *Cm. rotundifolium*; **Figure 3** and **Figure S4**), *Colchicum* (ca. 5 Ma between E African *Co. schimperianum* and S African *Co. albanense* and *Co. longipes*, **Figure 3** and **Figure S8**), the African clade of *Euphorbia* sect. *Esula* (ca. 7 Ma between S African and E African taxa; **Figure 3** and **Figure S10**), and *E.* sect. *Anthacanthae* (ca. 7.5 Ma separate subsects. *Platycephalae* and *Florispinae*; **Figure 3** and **Figure S11**).

Also from youngest to oldest, W-E disjunctions (**Figure 4**) can be found in the *Azorina* clade of *Campanula* (ca. 1 Ma between Cape Verdean *Ca. jacobaea* and Socotran *Ca. balfouri*; **Figure 3** and **Figure S3**), in *Hypericum* sect. *Campylosporus* (ca. 1.5 Ma within *H. quartinianum*; **Figure 3** and **Figure S13**), in *Aeonium* (1.7 Ma between E African *Ae. leucoblepharum* and a number of Macaronesian species; **Figure 3** and **Figure S2**), in *Cicer* (ca. 3.5 Ma between Canarian *Ci. canariense* and E African *Ci. cuneatum*; **Figure 3** and **Figure S7**), in *Adenocarpus* (ca. 4 Ma between E African *Ad. mannii* and a number of species in the *Ad. complicatus* complex; **Figure 3** and **Figure S1**), in *Euphorbia* sect. *Balsamis* (ca. 4 Ma between W African *Eu. balsamifera* subsp. *balsamifera* and E African-S Arabian *Eu. balsamifera* subsp. *adenensis*; **Figure 3** and **Figure S11**), in *Camptoloma* (ca. 5.5 Ma between Canarian *Cm. canariense*, on one hand, and E African *Cm. lyperiiflorum* and S African *Cm. rotundifolium*, on the other; **Figure 3** and **Figure S4**), *Eu.* sect. *Aphyllis* (ca. 5.5 Ma between Cape Verdean *Eu. tuckeyana* and all E African and S African species in this section; **Figure 3** and **Figure S9**), *Plocama* (ca. 6 Ma between Canarian *Pl. pendula* and S African *Pl. crocyllis* plus a number of E African/S Arabian *Plocama* species, **Figure 3** and **Figure S16**), in *Canarina* (6.5 Ma between Canarian *Cn. canariensis* and E African *Cn. eminii*; **Figure 3** and **Figure S6**), in *Kleinia* (ca. 7 Ma between the Macaronesian species, on one hand, and a clade of several E African species, on the other; **Figure 3** and **Figure S14**), in *Campylanthus* (ca. 7.5 Ma between the Macaronesian and the E African-S Arabian species in the genus; **Figure 3** and **Figure S5**), in *Geranium* subgen. *Robertium* (ca. 11 Ma between all E African species in this subgenus and a clade formed by W African taxa and a number of broadly distributed circum-Mediterranean and E Asian taxa; **Figure 3** and **Figure S12**), in the *Androsaemum* clade of *Hypericum* (ca. 17 Ma between Socotran *H. scopulorum*, *H. tortuosum* and Turkish *H. pamphylicum*, on one hand, and a number of Macaronesian and W Mediterranean species, on the other; **Figure 3** and **Figure S13**), and in *Sideroxylon* (ca. 17 Ma between Moroccan *S. spinosus* and E African *S. mascatense*; **Figure 3** and **Figure S16**).

Absolute Diversification Rates

Figure 5 and **Table S1** show results from net diversification rate analyses. Most lineages fall within the 95% CI of expected diversity under a no-extinction scenario ($\varepsilon = 0$) in the context of the RF family showing the highest rate of diversification (i.e., Asteraceae). However, some RF disjunct clades were significantly less diverse: W-E disjunctions in *Sideroxylon* (*S. spinosus* vs. *S. mascatense*), *Canarina* (*C. canariensis* vs. *C. eminii*), and *Hypericum* (*H. canariense* clade vs. *H. scopulorum* and *H. pamphylicum*). Other RF disjunct taxa were above the upper bound of the 95% CI: W-E(&S) disjunction in *Euphorbia* sect. *Aphyllis* (S), *Adenocarpus, Aeonium*, and *Campanula*; and E-S disjunction in *Plocama*. Otherwise, all taxa fell within the 95% CI with increasing ε values 0.5 and 0.9, except for *Sideroxylon*.

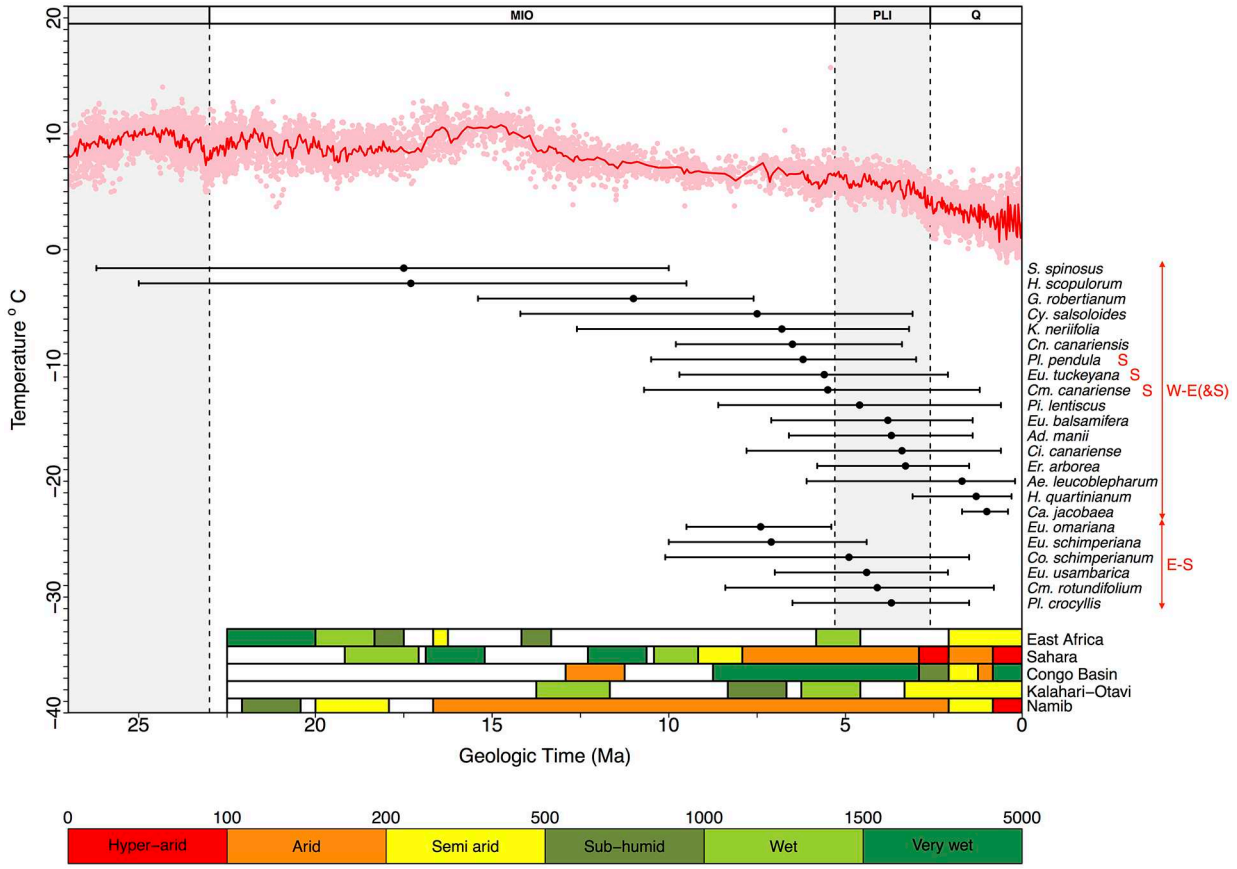


FIGURE 4 | Diagram showing estimated lineage divergence times (mean and 95% HPD confidence intervals) for Rand Flora disjunctions dated in this study and indicated in the MCC chronograms depicted in Figures S1–S17. W-E(&S): divergence times estimated between disjunct taxa distributed in Macaronesia-NW-W Africa vs. Eastern Africa (a red S indicates presence in Southern Africa); E-S: estimated divergence times between disjunct taxa distributed in southern Arabia-Eastern Africa vs. southern Africa. The red line above represents the change in global temperatures over the Cenozoic as reflected by global-deep-sea oxygen records compiled from Zachos et al. (2008); colored bars in the right bottom corner indicate climatic conditions in five regions that underwent major climate changes—either desertification or tropicalization—during the Neogene (adapted from Senut et al., 2009). Taxa names correspond to those in **Table 1**, plus two groups from the literature: *Pistacia lentiscus* and *Erica arborea* (see Discussion).

Interestingly these trends are generally repeated in the more encompassing lineages of the least diverse RF disjunct clades (e.g., *Canarina, Hypericum*, *Sideroxylon*). Notably, though *Camptoloma* has a low extant diversity given its age (three species diverging in the last 6 Myr), the subfamily it belongs to, that is Buddlejoideae, stands above the 95% CI for $\varepsilon = 0$ (**Figure 5**). Something similar can be observed in the case of *Kleinia*, which shows lower diversity than its encompassing lineage, tribe Senecioneae. Another example of potential diversification shift, though in the opposite direction, is that of *Euphorbia*, where the genus is significantly less diverse than expected given its age (for all ε values) but RF disjunct clades are species-richer than expected (i.e., *E.* sect. *Aphyllis*), except for those that fall within the 95% CI limits (e.g., *E.* sect. *Balsamis*, **Figure 5**).

When comparing crown vs. stem age it is noticeable that in some RF disjunct clades crown and stem ages are far apart: *Cicer canariensis* vs. *Ci. cuneatum* (crown age = 3.4 Ma, stem age = 12.2 Ma, with the stem age falling below the lower bound of 95% CIs when $\varepsilon = 0.0$ and 0.9; **Figure 5**). Other examples include, *Camptoloma* (crown age= 5.5 Ma, stem age = 10.2 Ma), *Campylanthus* (crown age = 7.5 Ma, stem age = 20.0 Ma), and most notably *Sideroxylon* (crown age = 17.4 Ma, stem age = 47.3 Ma, **Figure 5**).

Discussion

Rand Flora Disjunctions through Time

Engler's (1910) intuition on the *Tertiary* origins of the Afro-Macaronesian floristic element, aka Christ's (1910) Rand Flora, very much hit the mark on the timing of its assembly. Our divergence estimates for Rand Flora disjunctions span five successive time frames (**Figure 4**): Burdigalian, Tortonian, and Messinian Stages (within the Miocene), the Pliocene, and the Pleistocene. The two earliest disjunctions happen on genera *Sideroxylon* and *Hypericum* and date back to the Early Miocene (Burdigalian; 17.5 and 17.3 Ma, respectively), coinciding with the longest warming

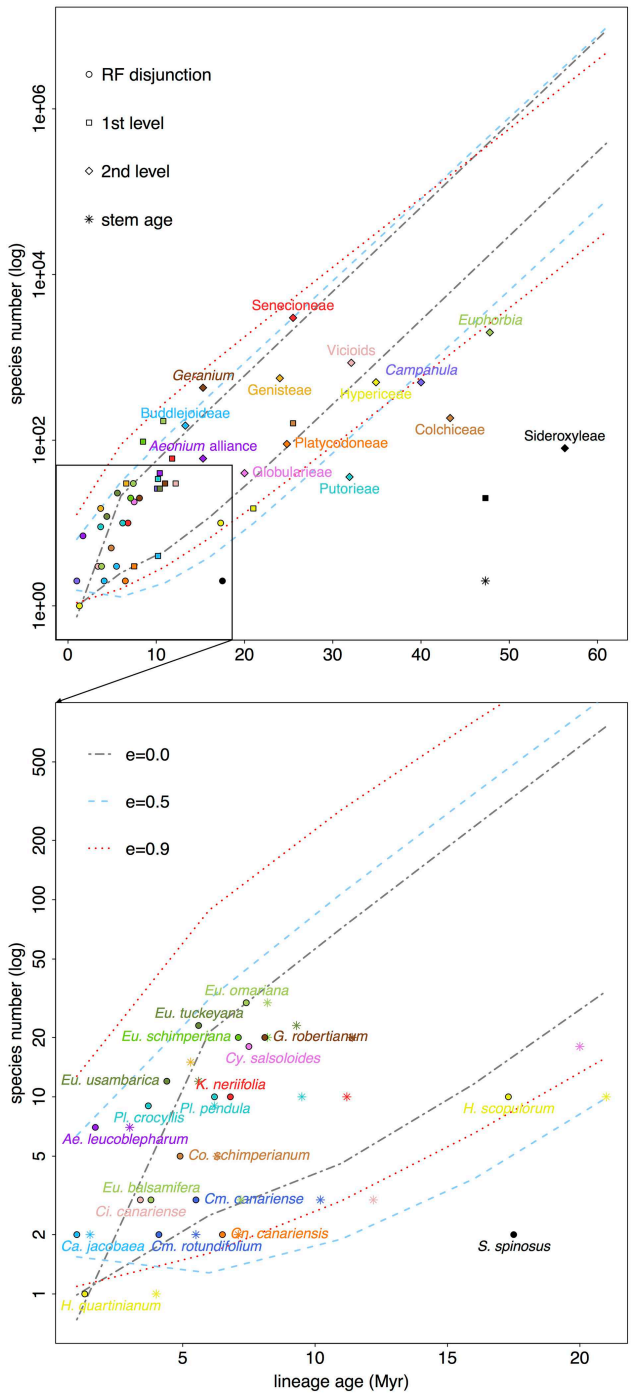


FIGURE 5 | RF lineages (names as in Table 1) are plotted according to their standing diversity (n) and age of the node (circle, crown; star, stem) corresponding to their disjunction (below). Successive encompassing lineages (above) also plotted (squares indicate the clade, section, subgenus the RF disjunct clade falls in; diamonds go one level above indicating genus, tribe, subfamily). Ninety five percent confidence intervals show expected diversity through time for a RF lineage that diversifies at the highest rate estimated (i.e., Asteraceae) given three possible scenarios: no extinction ($\varepsilon = 0$), turnover at equilibrium ($\varepsilon = 0.5$), and high extinction ($\varepsilon = 0.9$). See **Table S1** for associated net diversification rate estimates.

period of the Miocene (the Miocene Climatic Optimum; Zachos et al., 2008) and with the start of desertification in south-central Africa (Senut et al., 2009). Couvreur et al. (2008) also dated divergences in Annonaceae back to this time period and explained them in terms of a once-continuous Early Miocene rainforest that became fragmented by decreasing moisture brought by the closure of the Tethys Sea. The fact that *Sideroxylon* and *Hypericum* exhibit less xeric affinities than other RF lineages, and that their crown diversification dates back to the Paleogene (Meseguer et al., 2013; Stride et al., 2014), suggests these taxa could be relicts of an earlier megathermal flora (sensu Morley, 2000, 2003).

The next disjunction is that of *Geranium* subgen. *Robertium* and it dates back to the Late Miocene (Tortonian, 11.0 Ma). This disjunction follows a drastic decline in global temperatures (Late Miocene cooling, 11.6–5.3 Ma; Beerling et al., 2012) and coincides with the temporary closing of the Panama isthmus in America and a moist "washhouse" climate period in Europe (Böhme et al., 2008). This disjunction marks the separation of Macaronesian (e.g., *G. maderense*) and circum-Mediterranean taxa (e.g., *G. robertianum*), on one side, and E African species (e.g., *G. mascatense*), on the other, leaving open the possibility of a colonization of Macaronesia by a Mediterranean ancestor (**Figure 4** and **Figure S12**). Since the disjunction in *Geranium* subgen. *Robertium* is linked to a more humid period, rather than an increase on aridity, and because the possible Mediterranean origin of its Macaronesian taxa, this lineage does not exactly match the RF pattern.

Most other Neogene disjunctions seem to concentrate around the Miocene-Pliocene border (**Figure 4**). Messinian disjunctions can be observed in *Camptoloma, Campylanthus*, *Canarina*, *Euphorbia* sects. *Anthacanthae* and *Aphyllis*, *Kleinia*, and *Plocama*. Pliocene disjunctions are found in *Adenocarpus*, *Camptoloma*, *Cicer*, *Colchicum*, *Euphorbia*. sects. *Balsamis* and *Aphyllis*, and *Plocama*. These disjunctions follow two different geographic splits, W-E(&S) Africa and E-S Africa. W-E(&S) disjunctions present the widest temporal (as well as spatial) range. Besides the lineages dated here, other examples can be found in the literature of this W-E(&S) disjunction, e.g., according to Xie et al. (2014), in the Anacardiaceae *Pistacia lentiscus* and *P. aethiopica* diverged 4.55 Ma (see **Figure S17**). E-S disjunctions link South Africa and adjacent areas to the East African Rift Mountains, the Ethiopian Highlands, and the Arabian Peninsula. The timing of these E-S disjunctions (Mio-Pliocene) matches the uplift of the Eastern Arc Mountains (Sepulchre et al., 2006). The absence of W-S disjunctions is notable and probably results from African aridification having started in the early Miocene (some 17–16 Ma) in the region where the current Namib Desert stands. This aridification not only persisted through time in this area but also intensified and resulted in the formation of the Kalahari Desert (Senut et al., 2009), effectively limiting range expansions in this direction (W-S), in the absence of successful colonization following LDD. Even in the case of genus *Colchicum* (**Figure S8**), were S African species appear closely related to NW African ones, W Mediterranean species are always sister to E Mediterranean ones. These leaves open the possibility of a colonization of NW Africa (from S Africa) via E Africa and W Mediterranean populations

with subsequent extinction in E Africa. An alternative colonization from Central-West Asia into South Africa and NW Africa seems unlikely given the phylogeny of this genus (**Figure S8**), though proper biogeographic inference to test either possibility remains to be done. Indeed, Sanmartín et al. (2010) found a higher frequency of biotic exchange between NW-E African elements than with either E-S African or W-S African ones, where the latter elements were hardly connected, if at all, confirming our observations. We further argue that the magnitude of observed biotic exchange follows the history of desertification in Africa.

In all, the sequential timing of Neogene disjunctions in RF lineages, which is nonetheless concentrated in certain time intervals (e.g., Late Miocene-Pliocene), is in agreement with a scenario of range expansions (dispersal) in favorable times (windows of opportunity) and range contractions (extinction) as aridification flared up. Extinction results in absence (of a population, species, clade, or lineage) and thus leaves hard to track traces in phylogenies in the absence of fossil data (Meseguer et al., 2015). If repeated cycles of speciation, dispersal, and extinction take place in the same area over time, only taxa that optimize any (or a combination) of these processes (e.g., increased speciation, higher dispersal, lower extinction rates) will persist. It is to be expected that more recent populations, species, clades, or lineages show traces of these processes when compared to ancient ones.

On the other hand, our net diversification rate estimates (**Figure 5**) do no fully support an extinction explanation since, in the context of the family with the highest diversification rate among RF lineages, i.e., Asteraceae, most of the taxa fall inside the 95% CI under a no-extinction scenario ($\varepsilon = 0.0$). However, the method chosen to estimate net diversification rates (Magallón and Sanderson, 2001), though more appropriate given phylogeny size and sampling effort, is still limited. Crown diversification rates cannot be estimated for clades with 2 terminal taxa (see Materials and Methods), which is the case for several RF lineages (e.g., *Sideroxylon*). Additionally, the "method-of-moments" estimator performs well detecting declining diversity for old groups in exceedingly species-poor clades (Magallón and Sanderson, 2001; Warren and Hawkins, 2006) or young groups notably species-rich (recent radiations, Magallón and Sanderson, 2001), but we observed that statistical power is low to detect declines in diversity for young species-poor groups (e.g., *Camptoloma*). Most RF disjunct clades dated comprise less than 10 species—e.g., *Aeonium*, *Campanula*, *Camptoloma*, *Cicer*, *Colchicum*, *Euphorbia* sect. *Balsamis*, *Kleinia*, and *Plocama*—, limiting our ability to effectively detect the effects of extinction.

Nonetheless, if we focus on crown ages, disjunct clades in *Canarina, Hypericum*, and *Sideroxylon* are less diverse than expected, and given that their encompassing lineages (**Table 1**, **Figure 5**) also follow this trend, it would be safe to assume these lineages have indeed experienced high levels of extinction through time. Likewise, if we were to focus on stem ages, a few other groups fall below the no-extinction scenario ($\varepsilon = 0.0$), notably, *Camptoloma*, *Campylanthus*, and *Cicer*. Moreover, these groups exhibit wide-spanning (often >10 Ma) stem-crown intervals (see *Sideroxylon* or *Cicer* in **Figure 5**), an observation that has been tied to historically high extinction rates in recent diversification studies (Antonelli and Sanmartín, 2011; Nagalingum et al., 2011). This would further support the hypothesis that lower diversification rates in RF lineages could be explained in terms of increased extinction rather than a decrease in speciation rates.

Additionally, and given the aforementioned limitations of our diversification method of choice, it would also be safe to conclude that, within *Euphorbia,* sects. *Anthacanthae* (sect. *Balsamis* included), sect. *Esula*, and sect. *Aphyllis*, present higher diversity than expected (above the CI for $\varepsilon = 0.0$ in all cases, and also above the CI for $\varepsilon = 0.5$ for the former two clades), which is exceptional in the context of the genus, since *Euphorbia* is significantly poorer than expected for all ε values. Horn et al. (2014) also detected increased diversification rates in these sections of *Euphorbia*. Desertification-tropicalization cycles in Africa (Senut et al., 2009) suggest repeated reconnections between now disjunct RF regions since the Neogene, which would have permitted biotic exchange in favorable periods, whereas the isolation of these regions at unfavorable times would have induced speciation through vicariance, enhancing endemicity in these sub-humid/sub-xeric lineages. Molecular dating in tropical trees from the genus *Acridocapus* (Malpighiaceae; Davis et al., 2002) and the Annonaceae family (Couvreur et al., 2008) shows a similar pattern of connection phases between East African and Guineo-Congolian rainforest regions since the Oligocene following major climate shifts.

The youngest disjunctions, those of *Aeonium*, *Campanula*, and *Hypericum* sect. *Campylosporus,* are Pleistocene in age (**Figure 4**) and far too recent to result from the Neogene aridification of the African continent. Either rare LDD (i.e., *Aeonium*; Kim et al., 2008) or stepping-stone dispersal events (i.e., *Campanula*, Alarcón et al., pers. comm.), perhaps favored by Pleistocene cool and drier glacial cycles, could explain these more recent disjunct geographic patterns, as previously observed in other African taxa, e.g., *Convolvulus* (Carine, 2005), *Moraea* (Galley et al., 2007), or the tree heath (*Erica arborea*). Désamoré et al. (2011) took notice of successive range expansions of *Er. arborea* from an Eastern African center of diversity toward Northwest Africa, Southwest Europe, and Macaronesia, first during the Late Pliocene (ca. 3 Ma; **Figure 4**) and subsequently in the Pleistocene (ca. 1 Ma).

Redefining the Rand Flora Pattern

In a recent review, Linder (2014) synthesized the individual histories of numerous African lineages by recognizing five different "floras," which he defined as "groups of clades, which: (a) are largely found in the same area, (b) have largely the same extra-African geographical affinities, (c) share a diversification history, and (d) have a common maximum age." The "Rand Flora" does not fit well this definition. This *flora* does group a number of lineages that share the same geographic range (even if discontinuous), but they have slightly different climatic tolerances, i.e., sub-humid to sub-xeric or xerophilic, and they do not necessarily share the same extra-African geographical affinities.

Some RF lineages fall within what Linder (2014) terms "tropic-montane flora" (e.g., *Hypericum*, *Canarina*), others within the "arid flora" (e.g., *Kleinia*, *Campylanthus*). Some RF lineages are better connected with the Mediterranean Region (e.g., *Adenocarpus*), others with Asia and the Indo-Pacific Region (e.g., *Plocama*). Moreover, RF taxa on either side of any given disjunction (i.e., W-E or E-S) do no longer share a "diversification history," though they do share the same fate as other RF lineages with similar distribution. In fact, the different ages estimated here for the various RF disjunctions agree well with what has been termed biogeographic *pseudocongruence* (Donoghue and Moore, 2003), a phenomenon whereby two or more lineages display the same biogeographic pattern but with different temporal origins (Sanmartín, 2014). What is shared by all RF lineages is the nature of the climatic (ecological) barriers separating the taxa at either side of any given disjunction: arid regions such as the Sahara, the Kalahari or the Namib deserts, or the tropical lowlands in Central Africa. The congruence between RF disjunction ages and successive major climatic events in Africa during the Neogene (**Figure 4**) suggest that the ongoing aridification of the continent (or the "tropicalization" of Central Africa) affected RF lineages according to their different physiological (climatic) tolerances: more sub-humid lineages diverged first (e.g., *Sideroxylon*), more xeric later (e.g., *Campylanthus*).

One point of contention in the literature has been the limits of the Rand Flora with respect to the "Arid Corridor" or "Arid Track" (hereafter AC), a path repeatedly connecting southwest to north-east arid regions in Africa (and henceforth to central and southwest Asia) first proposed by Winterbottom (1967) and later expanded by de Winter (1966, 1971) and Verdcourt (1969). Bellstedt et al. (2012) defined the AC pattern as the disjunction occurring between Southern Africa and Eastern African-Southern Arabian xeric floristic elements. Linder (2014) considered the RF as an expansion of the AC to the west, in agreement with Jürgens' (1997) view. However, we consider that the RF and AC patterns are different. AC elements have more xeric preferences than the sub-humid to sub-xeric ones exhibited by RF elements. AC elements often extend into deserts (e.g., Namib, Kalahari, Sahara)—see studies by Beier et al. (2004) on *Fagonia* (Zygophyllaceae), Bellstedt et al. (2012) on *Zygophyllum* (also Zygophyllaceae), Carlson et al. (2012) on *Scabiosa* (Dipsacaceae), or Bruyns et al. (2014) on Ceropegieae— and have broader, more continuous distributions, plus they tend to be younger in age (often Pleistocene, coincident with Quaternary glaciation cycles). Our understanding is that this younger xeric AC elements move in parallel to RF taxa webbing with them in areas favorable to either, and thus confusing their limits. Something similar could have happened with Afromontane elements migrating south to north as the Eastern African mountains rose through the Mio-Pliocene; these elements are not part of the RF (e.g., *Iris*, *Moraea*, Galley et al., 2007).

In this study, we have provided a temporal framework for the Rand Flora pattern and estimated net diversification rates for 17 RF lineages. Our results provide some support to the historical view of an ancient African flora, whose current disjunct distribution was probably modeled by the successive waves of aridification events that have affected the African continent starting in the Miocene, but whose origin predates the latest events of Pleistocene climate change. These patterns were probably formed by a combination of climate-driven extinction and vicariance within a formerly widespread distribution. Whether these lineages all had a continuous, never interrupted, distribution that occupied all the area that now lies in between the extremes of the disjunction, or they had a somewhat narrower distribution in the past and they expanded their range tracking their habitat across the landscape in response to changing climate (e.g., along a corridor), is difficult to say with the current evidence. Discerning between these hypotheses will require the integration of phylogenetic, biogeographic and ecological approaches to reconstruct the ancestral ranges and climatic preferences of ancestral lineages (Mairal et al., 2015; Meseguer et al., 2015). Compared to speciation, extinction has received far less attention in studies focusing on the assembly of tropical biotas. Disentangling extinction from other processes is particularly difficult because the biodiversity we observe today is only a small fraction of that of the past. The Rand Flora pattern might offer a prime study model to understand the effects of climate-driven extinction in the shaping of continent-wide biodiversity patterns.

Author Contributions

IS and LP conceived and designed the study. LP analyzed the data with help from IS, RR, and MM. LP and IS co-wrote the text, with contributions from MH, RR, MM, and AM. All authors contributed with data compilation, figure preparation, or text comments. MM has copyright of all plant pictures, except for *Cicer canariense*.

Acknowledgments

This study was funded by the Spanish Ministry of Economy and Competitiveness (MINECO): Project AFFLORA, CGL2012-40129-C02-01 to IS. LP was funded by CSIC postdoctoral contract within AFFLORA. MH was funded by CGL2012-40129-C02-02, the Research Council of Norway (203822/E40) and a Ramón y Cajal Fellowship (RYC2009-04537). RR was supported by a JAE-DOC postdoctoral fellowship (MINECO) and the European Social Fund. MM and VC were supported by MINECO FPI predoctoral fellowships (BES-2010-037261 and BES-2013-065389 respectively). We thank Virginia Valcárcel (Department of Biology, UAM, Spain) for help with data compilation and literature revision during the earlier stages of the project, Andrea Briega (Department of Ecology, UAH, Spain) for help with data compilation, and Manuel Gil for providing a *Cicer canariense* picture.

Supplementary Material

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fgene.2015.00154/abstract

Supplementary Materials include descriptions of study groups with references, **Table S1**, and **Figures S1–S17**.

Table S1 | Net diversification rates (bd.ms) for all RF disjunct clades and their encompassing lineages (bold = highest crown.p, red when $n \leq 2$) under three possible scenarios: no extinction ($\varepsilon = 0$), turnover at equilibrium ($\varepsilon = 0.5$), and high extinction ($\varepsilon = 0.9$). Probability (crown.p) of obtaining a clade with the same size and age as the RF disjunction, given the background diversification rate of the encompassing clade/s and at increasing extinction fractions (bold = highest crown.p, italics $p < 0.05$). Stem and Crown ages in Myr.

Figures S1–S17 | BEAST MCC chronograms showing mean estimates and 95% high posterior density (HPD) confidence intervals for those nodes receiving 50% support. Branch width is proportional to PP support. Red colored taxa indicate Eastern African provenance; Macaronesia/western African taxa and southern African taxa are colored in blue and green, respectively. Calibration points are indicated with stars; RF disjunctions within each lineage discussed in the text and represented in **Figures 3–5** are indicated with arrows.

References

Alarcón, M., Roquet, C., García-Fernández, A., Vargas, P., and Aldasoro, J. J. (2013). Phylogenetic and phylogeographic evidence for a Pleistocene disjunction between *Campanula jacobaea* (Cape Verde Islands) and *C. balfourii (Socotra)*. *Mol. Phylogenet. Evol.* 69, 828–836. doi: 10.1016/j.ympev.2013.06.021

Andrus, N., Trusty, J., Santos-Guerra, A., Jansen, R. K., and Francisco-Ortega, J. (2004). Using molecular phylogenies to test phytogeographical links between East/South Africa–Southern Arabia and the Macaronesian islands—a review, and the case of *Vierea* and *Pulicaria* section Vieraeopsis (Asteraceae). *Taxon* 53, 333–333. doi: 10.2307/4135612

Antonelli, A., and Sanmartín, I. (2011). Mass extinction, gradual cooling, or rapid radiation? Reconstructing the spatiotemporal evolution of the ancient angiosperm genus *Hedyosmum* (Chloranthaceae) using empirical and simulated approaches. *Syst. Biol.* 60, 596–615. doi: 10.1093/sysbio/syr062

Axelrod, D. I., and Raven, P. H. (1978). "Late Cretaceous and Tertiary vegetation history of Africa," in *Biogeography and Ecology of southern Africa*, ed M. J. A. Werger (The Hague: Dr W. Junk bv Publishers), 77–130.

Backlund, M., Bremer, B., and Thulin, M. (2007). Paraphyly of Paederieae, recognition of Putorieae and expansion of *Plocama* (Rubiaceae-Rubioideae). *Taxon* 56, 315–328. Available online at: http://www.ingentaconnect.com/content/iapt/tax/2007/00000056/00000002/art00006

Barreda, V. D., Palazzesi, L., Katinas, L., Crisci, J. V., Tellería, M. C., Bremer, K., et al. (2012). An extinct Eocene taxon of the daisy family (Asteraceae): evolutionary, ecological and biogeographical implications. *Ann. Bot.* 109, 127–134. doi: 10.1093/aob/mcr240

Barres, L., Sanmartín, I., Anderson, C. L., Susanna, A., Buerki, S., Galbany-Casals, M., et al. (2013). Reconstructing the evolution and biogeographic history of tribe *Cardueae* (Compositae). *Am. J. Bot.* 100, 867–882. doi: 10.3732/ajb.1200058

Barres, L., Vilatersana, R., Molero, J., Susanna, A., and Galbany-Casals, M. (2011). Molecular phylogeny of *Euphorbia* subg. *Esula sect. Aphyllis* (Euphorbiaceae) inferred from nrDNA and cpDNA markers with biogeographic insights. *Taxon* 60, 705–720. Available online at: http://www.ingentaconnect.com/content/iapt/tax/2011/00000060/00000003/art00007

Beerling, D. J., Taylor, L. L., Bradshaw, C. D., Lunt, D. J., Valdes, P. J., Banwart, S. A., et al. (2012). Ecosystem CO2 starvation and terrestrial silicate weathering: mechanisms and global-scale quantification during the late Miocene. *J. Ecol.* 100, 31–41. doi: 10.1111/j.1365-2745.2011.01905.x

Beier, B. A., Nylander, J. A. A., Chase, M. W., and Thulin, M. (2004). Phylogenetic relationships and biogeography of the desert plant genus *Fagonia* (Zygophyllaceae), inferred by parsimony and Bayesian model averaging. *Mol. Phylogenet. Evol.* 33, 91–108. doi: 10.1016/j.ympev.2004.05.010

Bell, C. D., Soltis, D. E., and Soltis, P. S. (2010). The age and diversification of the angiosperms re-revisited. *Am. J. Bot.* 97, 1296–1303. doi: 10.3732/ajb.0900346

Bellstedt, D. U., Galley, C., Pirie, M. D., and Linder, H. P. (2012). The migration of the palaeotropical arid flora: zygophylloideae as an example. *Syst. Bot.* 37, 951–959. doi: 10.1600/036364412X656608

Böhme, M., Ilg, A., and Winklhofer, M. (2008). Late Miocene "washhouse" climate in Europe. *Earth Planet. Sci. Lett.* 275, 393–401. doi: 10.1016/j.epsl.2008.09.011

Bonnefille, R. (2011). "Rainforest responses to past climate changes in tropical Africa," in *Tropical Rainforest Responses to Climate Change, 2nd Edn*, eds M. Bush, J. Flenley, and W. Gosling (Berlin; Heidelberg: Springer-Verlag), 125–184.

Bramwell, D. (1985). Contribución a la biogeografía de las Islas Canarias. *Bot. Macaronésica* 14, 3–34.

Bremer, B., and Eriksson, T. (2009). Time tree of Rubiaceae: phylogeny and dating the family, subfamilies, and tribes. *Int. J. Plant Sci.* 170, 766–793. doi: 10.1086/599077

Bruyns, P. V., Klak, C., and Hanáček, P. (2014). Evolution of the stapeliads (Apocynaceae–Asclepiadoideae)—repeated major radiation across Africa in an Old World group. *Mol. Phylogenet. Evol.* 77, 251–263. doi: 10.1016/j.ympev.2014.03.022

Carine, M. A. (2005). Spatio-temporal relationships of the Macaronesian endemic flora: a relictual series or window of opportunity? *Taxon* 54, 895–903. doi: 10.2307/25065476

Carlson, S. E., Linder, H. P., and Donoghue, M. J. (2012). The historical biogeography of *Scabiosa* (Dipsacaceae): implications for Old World plant disjunctions. *J. Biogeogr.* 39, 1086–1100. doi: 10.1111/j.1365-2699.2011.02669.x

Chacón, J., and Renner, S. S. (2014). Assessing model sensitivity in ancestral area reconstruction using LAGRANGE: a case study using the Colchicaceae family. *J. Biogeogr.* 41, 1414–1427. doi: 10.1111/jbi.12301

Christ, H. (1892). Exposé sur le rôle que joue dans le domaine de nos flores la flore dite *ancienne africaine*. *Arch. Sci. Phys. Nat. Genève* 3, 369–374.

Christ, H. (1910). *Die Geographie der Farne*. Jena: Verlag von Gustav Fischer. 1–358.

Coleman, M., Liston, A., Kadereit, J. W., and Abbott, R. J. (2003). Repeat intercontinental dispersal and Pleistocene speciation in disjunct Mediterranean and desert *Senecio* (Asteraceae). *Am. J. Bot.* 90, 1446–1454. doi: 10.3732/ajb.90.10.1446

Couvreur, T. L. (2015). Odd man out: why are there fewer plant species in African rain forests? *Plant Syst. Evol.* 301, 1299–1313. doi: 10.1007/s00606-014-1180-z

Couvreur, T. L., Chatrou, L. W., Sosef, M. S., and Richardson, J. E. (2008). Molecular phylogenetics reveal multiple tertiary vicariance origins of the African rain forest trees. *BMC Biol.* 6:54. doi: 10.1186/1741-7007-6-54

Cubas, P., Pardo, C., Tahiri, H., and Castroviejo, S. (2010). Phylogeny and evolutionary diversification of *Adenocarpus* DC. *(Leguminosae). Taxon* 59, 720–732. Available online at: http://www.ingentaconnect.com/content/iapt/tax/2010/00000059/00000003/art00005

Davis, C. C., Bell, C. D., Fritsch, P. W., and Mathews, S. (2002). Phylogeny of *Acridocarpus-Brachylophon* (Malpighiaceae): implications for Tertiary tropical floras and Afroasian biogeography. *Evolution* 56, 2395–2405. doi: 10.1111/j.0014-3820.2002.tb00165.x

del Hoyo, A., García-Marín, J. L., and Pedrola-Monfort, J. (2009). Temporal and spatial diversification of the African disjunct genus *Androcymbium* (Colchicaceae). *Mol. Phylogenet. Evol.* 53, 848–861. doi: 10.1016/j.ympev.2009.08.005

Désamoré, A., Laenen, B., Devos, N., Popp, M., González-Mancebo, J. M., Carine, M. A., et al. (2011). Out of Africa: north-westwards Pleistocene expansions of the heather *Erica arborea*. *J. Biogeogr.* 38, 164–176. doi: 10.1111/j.1365-2699.2010.02387.x

de Winter, B. (1966). Remarks on the distribution of some desert plants in Africa. *Palaeoecol. Afr.* 1, 188–189.

de Winter, B. (1971). Floristic relationships between the northern and southern arid areas in Africa. *Mitt. Bot. Staatssamml. Munch.* 10, 424–437.

Doláková, N., Kovaìcovaì, M., and Basistovaì, P. (2011). Badenian (Langhian-Early Serravallian) palynoflora for the Carpathian Foredeep and Vienna Basin (Czech and Slovak Republics). *Acta Ent. Mus. Nat. Pra.* 67, 63–71. Available online at: http://www.muni.cz/research/publications/955017

Donoghue, M. J., and Moore, B. R. (2003). Toward an integrative historical biogeography. *Integr. Comp. Biol.* 43, 261–270. doi: 10.1093/icb/43.2.261

Drummond, A. J., Ho, S. Y., Phillips, M. J., and Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biol.* 4:e88. doi: 10.1371/journal.pbio.0040088

Drummond, A. J., Suchard, M. A., Xie, D., and Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29, 1969–1973. doi: 10.1093/molbev/mss075

Engler, A. (1879). "Die extratropischen Gebiete der nördlichen Hemisphäre. Mit einer chromolithographischen Karte," in *Versuch einer Entwicklungsgeschichte der Pflanzenwelt: insbesondere der Florengebiete seit der Tertiärperiode. Theil 1*, ed A. Engler (Leipzig: Verlag von Wilhelm Engelmann), 1–202.

Engler, A. (1910). "Die Pflanzenwelt Afrikas inbesondere seiner tropischen Gebiete. Gründzuge der Pflanzenverbreitung in Afrika un die Characterpflanzen Afrikas," in *Die Vegetation der Erde*, eds A. Engler and O. Drude (Leipzig: Verlag von Wilhelm Engelmann), 1030.

Fischer, A. G. (1960). Latitudinal variations in organic diversity. *Evolution* 14, 64–81. doi: 10.2307/2405923

Fiz, O., Vargas, P., Alarcón, M., Aedo, C., García, J. L., and Aldasoro, J. J. (2008). Phylogeny and historical biogeography of Geraniaceae in relation to climate changes and pollination ecology. *Syst. Bot.* 33, 326–342. doi: 10.1600/036364408784571482

Forster, J. R. (1778). *Observations Made During a Voyage Round the World, on Physical Geography, Natural History, and Ethic Philosophy*. London: G. Robinson.

Galley, C., Bytebier, B., Bellstedt, D. U., and Linder, H. P. (2007). The Cape element in the Afrotemperate flora: from Cape to Cairo? *Proc. R. Soc. B* 274, 535–543. doi: 10.1098/rspb.2006.0046

Harmon, L. J., Weir, J. T., Brock, C. D., Glor, R. E., and Challenger, W. (2008). GEIGER: investigating evolutionary radiations. *Bioinformatics* 24, 129–131. doi: 10.1093/bioinformatics/btm538

Ho, S. Y., and Phillips, M. J. (2009). Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Syst. Biol.* 58, 367–380. doi: 10.1093/sysbio/syp035

Horn, J. W., Xi, Z., Riina, R., Peirson, J. A., Yang, Y., Dorsey, B. L., et al. (2014). Evolutionary burst in *Euphorbia* (Euphorbiaceae) are linked with Photosynthetic pathway. *Evolution* 68, 3485–3504. doi: 10.1111/evo.12534

Jablonski, D., Roy, K., and Valentine, J. W. (2006). Out of the tropics: evolutionary dynamics of the latitudinal diversity gradient. *Science* 314, 102–106. doi: 10.1126/science.1130880

Jacobs, B. F., Pan, A. D., and Scotese, C. R. (2010). "A review of the Cenozoic vegetation history of Africa," in *Cenozoic Mammals of Africa*, eds L. Werdelin and W. J. Sanders (Berkeley, CA: University of California Press), 57–72.

Javadi, F., Wojciechowski, M. F., and Yamaguchi, H. (2007). Geographical diversification of the genus *Cicer* (Leguminosae: Papilionoideae) inferred from molecular phylogenetic analyses of chloroplast and nuclear DNA sequences. *Bot. J. Linn. Soc.* 154, 175–186. doi: 10.1111/j.1095-8339.2007.00649.x

Jürgens, N. (1997). Floristic biodiversity and history of African arid regions. *Biodiv. Conserv.* 6, 495–514. doi: 10.1023/A:1018325026863

Kim, S. C., McGowen, M. R., Lubinsky, P., Barber, J. C., Mort, M. E., and Santos-Guerra, A. (2008). Timing and tempo of early and successive adaptive radiations in Macaronesia. *PLoS ONE* 3:e2139. doi: 10.1371/journal.pone.0002139

Kornhall, P., Heidari, N., and Bremer, B. (2001). Selagineae and Manuleeae, two tribes or one? Phylogenetic studies in the Scrophulariaceae. *Plant Syst. Evol.* 228, 199–218. doi: 10.1007/s006060170029

Lavin, M., Herendeen, P. S., and Wojciechowski, M. F. (2005). Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the Tertiary. *Syst. Biol.* 54, 575–594. doi: 10.1080/10635150590947131

Lavin, M., Wojciechowski, M. F., Richman, A., Rotella, J., Sanderson, M. J., and Matos, A. B. (2001). Identifying Tertiary radiations of Fabaceae in the Greater Antilles: alternatives to cladistic vicariance analysis. *Int. J. Plant Sci.* 162, S53–S76. doi: 10.1086/323474

Lebrun, J. (1947). "Essai sur l'origine et le développement de la flore," in *Exploration du Parc National Albert. Mission J. Lebrun (1937–1938). La végétation de la plaine alluviale au sud du Lac Édouard*, ed J. Lebrun (Brussels: Inst. des Parcs Nationaux du Congo Belge). Fasc. 1. 2ème Part., 115–397.

Lebrun, J. (1961). Les deux flores d'Afrique tropicale. *Acad. Roy. Belg. Cl. Sci. Mém. (coll. 8. 2ème sér.)* 32, 1–82.

Linder, H. P. (2014). The evolution of African plant diversity. *Front. Ecol. Evol.* 2, 1–14. doi: 10.3389/fevo.2014.00038

Lomolino, M. V., Riddle, B. R., Whittaker, R. J., and Brown, J. H. (2010). *Biogeography. 4th Edn*. Sunderland, MA: Sinauer Associates, Inc. 764.

Maclean, I. M. D., and Wilson, R. J. (2011). Recent ecological responses to climate change support predictions of high extinction risk. *Proc. Natl. Acad. Sci. U.S.A.* 108, 12337–12342. doi: 10.1073/pnas.1017352108

Maddison, W. P., and Maddison, D. R. (2014). *Mesquite: A Modular System for Evolutionary Analysis. Version 3.01*. Available online at: http://mesquiteproject.org

Magallón, S., and Sanderson, M. J. (2001). Absolute diversification rates in angiosperm clades. *Evolution* 55, 1762–1780. doi: 10.1111/j.0014-3820.2001.tb00826.x

Mairal, M., Pokorny, L., Aldasoro, J. J., Alarcón, M., and Sanmartín, I. (2015). Ancient vicariance and climate-driven extinction explain continental-wide disjunctions in Africa: the case of the Rand Flora genus *Canarina* (Campanulaceae). *Mol. Ecol.* 24, 1335–1354. doi: 10.1111/mec.13114

Maley, J. (1996). The African rain forest—main characteristics of changes in vegetation and climate from the Upper Cretaceous to the Quaternary. *Proc. Roy. Soc. Edinburgh Sect. B. Biol. Sci.* 104, 31–73. doi: 10.1017/S0269727000006114

Maley, J. (2000). Last Glacial Maximum lacustrine and fluviatile formations in the Tibesti and other Saharan mountains, and large-scale climatic teleconnections linked to the activity of the Subtropical Jet Stream. *Glob. Planet. Change* 26, 121–136. doi: 10.1016/S0921-8181(00)00039-4

Manning, J., Forest, F., and Vinnersten, A. (2007). The genus *Colchicum* L. redefined to include Androcymbium Willd. based on molecular evidence. *Taxon* 56, 872–882. doi: 10.2307/25065868

Meseguer, A. S., Aldasoro, J. J., and Sanmartín, I. (2013). Bayesian inference of phylogeny, morphology and range evolution reveals a complex evolutionary history in St. John's wort (Hypericum). *Mol. Phylogenet. Evol.* 67, 379–403. doi: 10.1016/j.ympev.2013.02.007

Meseguer, A. S., Lobo, J. M., Ree, R., Beerling, D. J., and Sanmartín, I. (2015). Integrating Fossils, Phylogenies, and Niche Models into Biogeography to reveal ancient evolutionary history: the Case of *Hypericum* (Hypericaceae). *Syst. Biol.* 64, 215–232. doi: 10.1093/sysbio/syu088

Monod, T. (1971). Remarques sur les symétries floristiques des zones sèches nord et sud en Afrique. *Mitt. Bot. Staatssamml. München* 10, 375–423.

Morley, R. J. (2000). *Origin and Evolution of Tropical Rain Forests*. Chinchester: John Wiley and Sons Ltd. 1–362.

Morley, R. J. (2003). Interplate dispersal paths for megathermal angiosperms. *Perspect. Plant Ecol. Evol. Syst.* 6, 5–20. doi: 10.1078/1433-8319-00039

Morlon, H. (2014). Phylogenetic approaches for studying diversification. *Ecol. Lett.* 17, 508–525. doi: 10.1111/ele.12251

Mort, M. E., Soltis, D. E., Soltis, P. S., Francisco-Ortega, J., and Santos-Guerra, A. (2002). Phylogenetics and evolution of the Macaronesian clade of Crassulaceae inferred from nuclear and chloroplast sequence data. *Syst. Bot.* 27, 271–288. doi: 10.1043/0363-6445-27.2.271

Mort, M. E., Soltis, D. E., Soltis, P. S., Santos-Guerra, A., and Francisco-Ortega, J. (2007). Physiological evolution and association between physiology and growth form in *Aeonium* (Crassulaceae). *Taxon* 56, 453–464.

Nagalingum, N. S., Marshall, C. R., Quental, T. B., Rai, H. S., Little, D. P., and Mathews, S. (2011). Recent synchronous radiation of a living fossil. *Science* 334, 796–799. doi: 10.1126/science.1209926

Nagy, E. (1963). Some new spore and pollen species from the Neogene of the Mecsek Mountain. *Acta Bot. Hung.* 9, 387–404.

Nathan, R. (2006). Long-distance dispersal of plants. *Science* 313, 786–788. doi: 10.1126/science.1124975

Navarro-Pérez, M. L., López, J., Fernández-Mazuecos, M., Rodríguez-Riaño, T., Vargas, P., and Ortega-Olivencia, A. (2013). The role of birds and insects in pollination shifts of *Scrophularia* (Scrophulariaceae). *Mol. Phylogenet. Evol.* 69, 239–254. doi: 10.1016/j.ympev.2013.05.027

Nylander, J. A. A. (2004). *MrModeltest v2.2 Program Distributed by the Author*. Uppsala: Evolutionary biology centre, Uppsala University, 2.

Osborne, A. H., Vance, D., Rohling, E. J., Barton, N., Rogerson, M., and Fello, N. (2008). A humid corridor across the Sahara for the migration of early modern humans out of Africa 120,000 years ago. *Proc. Natl. Acad. Sci. U.S.A.* 105, 16444–16447. doi: 10.1073/pnas.0804472105

Oxelman, B., Kornhall, P., Olmstead, R. G., and Bremer, B. (2005). Further disintegration of Scrophulariaceae. *Taxon* 54, 411–425. doi: 10.2307/25065369

Peirson, J. A., Bruyns, P. V., Riina, R., Morawetz, J. J., and Berry, P. E. (2013). A molecular phylogeny and classification of the largely succulent and mainly African *Euphorbia* subg. *Athymalus (Euphorbiaceae)*. *Taxon* 62, 1178–1199. doi: 10.12705/626.12

Pelser, P. B., Abbott, R. J., Comes, H. P., Milton, J. J., Moeller, M., Looseley, M. E., et al. (2012). The genetic ghost of an invasion past: colonization and extinction revealed by historical hybridization in *Senecio*. *Mol. Ecol.* 21, 369–387. doi: 10.1111/j.1365-294X.2011.05399.x

Pelser, P. B., Kennedy, A. H., Tepe, E. J., Shidler, J. B., Nordenstam, B., Kadereit, J. W., et al. (2010). Patterns and causes of incongruence between plastid and nuclear Senecioneae (Asteraceae) phylogenies. *Am. J. Bot.* 97, 856–873. doi: 10.3732/ajb.0900287

Pelser, P. B., Nordenstam, B., Kadereit, J. W., and Watson, L. E. (2007). An ITS phylogeny of tribe Senecioneae (Asteraceae) and a new delimitation of Senecio L. *Taxon* 56, 1077–1077. doi: 10.2307/25065905

Pennington, R. T., Cronk, Q. C., and Richardson, J. A. (2004). Introduction and synthesis: plant phylogeny and the origin of major biomes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 359, 1455–1464. doi: 10.1098/rstb.2004.1539

Pennington, R. T., Lavin, M., Särkinen, T., Lewis, G. P., Klitgaard, B. B., and Hughes, C. E. (2010). Contrasting plant diversification histories within the Andean biodiversity hotspot. *Proc. Natl. Acad. Sci. U.S.A.* 107, 13783–13787. doi: 10.1073/pnas.1001317107

Plana, V. (2004). Mechanisms and tempo of evolution in the African Guineo–Congolian rainforest. *Philos. Trans. R. Soc. Lond. B, Biol. Sci.* 359, 1585–1594. doi: 10.1098/rstb.2004.1535

Qian, H., and Ricklefs, R. E. (2000). Large-scale processes and the Asian bias in species diversity of temperate plants. *Nature* 407, 180–182. doi: 10.1038/35025052

Quézel, P. (1978). Analysis of the flora of mediterranean and saharan africa. *Ann. Missouri Bot. Gard.* 65, 479–534. doi: 10.2307/2398860

Rabosky, D. L., Grundler, M., Anderson, C., Shi, J. J., Brown, J. W., Huang, H., et al. (2014). BAMMtools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods Ecol. Evol.* 5, 701–707. doi: 10.1111/2041-210X.12199

Rambaut, A. (2002). *Se-Al, Version 2.0 a11. Computer Program Distributed by the Author.* Available online at: http://tree.bio.ed.ac.uk/software/seal/

Rambaut, A. (2009). *FigTree, Version 1.4.2. Computer Program Distributed by the Author.* http://tree.bio.ed.ac.uk/software/figtree/

Rambaut, A., Suchard, M. A., Xie, W., and Drummond, A.J. (2013). *Tracer v1.6.* Available online at: http://tree.bio.ed.ac.uk/soft-ware/tracer

Raven, P. H., and Axelrod, D. I. (1974). Angiosperm biogeography and past continental movements. *Ann. Missouri Bot. Gard.* 61, 539–673. doi: 10.2307/2395021

Richards, P. W. (1973). "Africa, the 'Odd man out,'" in *Tropical Forest Ecosystems of Africa and South America: A Comparative Review*, eds B. J. Meggers, E. S. Ayensu, and W. D. Duckworth (Washington, DC: Smithsonian Institution Press), 21–26.

Ricklefs, R. E. (2007). Estimating diversification rates from phylogenetic information. *Trends Ecol. Evol.* 22, 601–610. doi: 10.1016/j.tree.2007.06.013

Riina, R., Peirson, J. A., Geltman, D. V., Molero, J., Frajman, B., Pahlevani, A., et al. (2013). A worldwide molecular phylogeny and classification of the leafy spurges, *Euphorbia* subgenus *Esula* (Euphorbiaceae). *Taxon* 62, 316–342. doi: 10.12705/622.3

Roquet, C., Sanmartín, I., García-Jacas, N., Sáez, L., Susanna, A., Wikström, N., et al. (2009). Reconstructing the history of Campanulaceae with a Bayesian approach to molecular dating and dispersal–vicariance analyses. *Mol. Phylogenet. Evol.* 52, 575–587. doi: 10.1016/j.ympev.2009.05.014

Sanmartín, I. (2014). "Biogeography," in *The Tree of Life*, eds P. Vargas and R. Zardoya (Sunderland, MA: Sinauer Associates, Inc.), 555–576.

Sanmartín, I., Anderson, C. L., Alarcon, M., Ronquist, F., and Aldasoro, J. J. (2010). Bayesian island biogeography in a continental setting: the Rand Flora case. *Biol. Lett.* 6, 703–707. doi: 10.1098/rsbl.2010.0095

Senut, B., Pickford, M., and Ségalen, L. (2009). Neogene desertification of Africa. *C. R. Geoscience* 341, 591–602. doi: 10.1016/j.crte.2009.03.008

Sepulchre, P., Ramstein, G., Fluteau, F., Schuster, M., Tiercelin, J. J., and Brunet, M. (2006). Tectonic uplift and Eastern Africa aridification. *Science* 313, 1419–1423. doi: 10.1126/science.1129158

Simmons, M. P. (2004). Independence of alignment and tree search. *Mol. Phylogenet. Evol.* 31, 874–879. doi: 10.1016/j.ympev.2003.10.008

Smedmark, J. E., and Anderberg, A. A. (2007). Boreotropical migration explains hybridization between geographically distant lineages in the pantropical clade Sideroxyleae (Sapotaceae). *Am. J. Bot.* 94, 1491–1505. doi: 10.3732/ajb.94.9.1491

Smedmark, J. E., Swenson, U., and Anderberg, A. A. (2006). Accounting for variation of substitution rates through time in Bayesian phylogeny reconstruction of Sapotoideae (Sapotaceae). *Mol. Phylogenet. Evol.* 39, 706–721. doi: 10.1016/j.ympev.2006.01.018

Stadler, T. (2013). Recovering speciation and extinction dynamics based on phylogenies. *J. Evol. Biol.* 26, 1203–1219. doi: 10.1111/jeb.12139

Stevens, G. C. (1989). The latitudinal gradient in geographical range: how so many species coexist in the tropics. *Am. Nat.* 133, 240–256. doi: 10.1086/284913

Stride, G., Nylinder, S., and Swenson, U. (2014). Revisiting the biogeography of *Sideroxylon* (Sapotaceae) and an evaluation of the taxonomic status of *Argania* and *Spiniluma*. *Austral. Syst. Bot.* 27, 104–118. doi: 10.1071/SB14010

Swofford, D. (2002). *Phylogenetic Analysis Using Parsimony (PAUP* v4.0b)*. Sunderland, MA: Sinauer Associates.

Thiv, M., Thulin, M., Hjertson, M., Kropf, M., and Linder, H. P. (2010). Evidence for a vicariant origin of Macaronesian–Eritreo/Arabian disjunctions in *Campylanthus* Roth (Plantaginaceae). *Mol. Phylogenet. Evol.* 54, 607–616. doi: 10.1016/j.ympev.2009.10.009

Trauth, M. H., Larrasoaña, J. C., and Mudelsee, M. (2009). Trends, rhythms and events in Plio-Pleistocene African climate. *Quat. Sci. Rev.* 28, 399–411. doi: 10.1016/j.quascirev.2008.11.003

Van Campo, E. (1989). Flore pollinique du Miocene superieur de Venta del Moro (Espagne). *Acta Palynol.* 1, 9–32.

van Zinderen Bakker, E. M. Sr. (1978). "Quaternary vegetation changes in southern Africa," in *Biogeography and Ecology of southern Africa*, ed M. J. A. Werger (The Hague: Dr W. Junk bv Publishers), 131–143.

Verdcourt, B. (1969). The arid corridor between the northeast and southwest areas of Africa. *Palaeoecol. Afr.* 4, 140–144.

von Humboldt, A., and Bonpland, A. (1805). *Essai sur la Geìographie des Plantes; Accompagneì d'un Tableau Physique des Reìgions Eìquinoxiales, Fondeì sur des Mesures exeìcuteìes, Depuis le DixieÈme Degreì de Latitude Boreìale Jusqu'au DixieÈme Degreì de Latitude Australe, Pendant les Anneìes 1799, 1800, 1801, 1802 et 1803*. Paris: chez Levrault, Schoell et compagnie, libraries, 155.

Wallace, A. R. (1878). *Tropical Nature, and Other Essays*. London: Macmillan and co. 356.

Warren, B. H., and Hawkins, J. A. (2006). The distribution of species diversity across a flora's component lineages: dating the Cape's 'relicts'. *Proc. R. Soc. B* 273, 2149–2158. doi: 10.1098/rspb.2006.3560

Wheeler, T. J., and Kececioglu, J. D. (2007). Multiple alignment by aligning alignments. *Bioinformatics* 23, i559–i568. doi: 10.1093/bioinformatics/btm226

Wiens, J. J. (2004). What is speciation and how should we study it? *Am. Nat.* 163, 914–922. doi: 10.1086/386552

Wiens, J. J. (2011). The causes of species richness patterns across space, time, and clades and the role of "ecological limits." *Q. Rev. Biol.* 86, 75–96. doi: 10.1086/659883

Wiens, J. J., and Donoghue, M. J. (2004). Historical biogeography, ecology and species richness. *Trends Ecol. Evol.* 19, 639–644. doi: 10.1016/j.tree.2004.09.011

Winterbottom, J. M. (1967). Climatological implications of avifaunal resemblances between South Western Africa and Somaliland. *Palaeoecol. Afr.* 2, 77–79.

Xie, L., Yang, Z. Y., Wen, J., Li, D. Z., and Yi, T. S. (2014). Biogeographic history of *Pistacia* (Anacardiaceae), emphasizing the evolution of the Madrean-Tethyan and the eastern Asian-Tethyan disjunctions. *Mol. Phylogenet. Evol.* 77, 136–146. doi: 10.1016/j.ympev.2014.04.006

Zachos, J. C., Dickens, G. R., and Zeebe, R. E. (2008). An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature* 451, 279–283. doi: 10.1038/nature06588

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Pokorny, Riina, Mairal, Meseguer, Culshaw, Cendoya, Serrano, Carbajal, Ortiz, Heuertz and Sanmartín. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

View publication stats

MOLECULAR ECOLOGY

Molecular Ecology (2015) doi: 10.1111/mec.13282

Palaeo-islands as refugia and sources of genetic diversity within volcanic archipelagos: the case of the widespread endemic *Canarina canariensis* (Campanulaceae)

M. MAIRAL,* I. SANMARTÍN,*[1] J. J. ALDASORO,† V. CULSHAW,* I. MANOLOPOULOU‡ and M. ALARCÓN†[1]

Real Jardín Botánico (RJB-CSIC), 28014 Madrid, Spain, †Institut Botànic de Barcelona (IBB-CSIC-ICUB), 08038 Barcelona, Spain, ‡University College London, WC1E 6BT London, UK

Abstract

Geographical isolation by oceanic barriers and climatic stability has been postulated as some of the main factors driving diversification within volcanic archipelagos. However, few studies have focused on the effect that catastrophic volcanic events have had on patterns of within-island differentiation in geological time. This study employed data from the chloroplast (cpDNA haplotypes) and the nuclear (AFLPs) genomes to examine the patterns of genetic variation in *Canarina canariensis*, an iconic plant species associated with the endemic laurel forest of the Canary Islands. We found a strong geographical population structure, with a first divergence around 0.8 Ma that has Tenerife as its central axis and divides Canarian populations into eastern and western clades. Genetic diversity was greatest in the geologically stable 'palaeo-islands' of Anaga, Teno and Roque del Conde; these areas were also inferred as the ancestral location of migrant alleles towards other disturbed areas within Tenerife or the nearby islands using a Bayesian approach to phylogeographical clustering. Oceanic barriers, in contrast, appear to have played a lesser role in structuring genetic variation, with intra-island levels of genetic diversity larger than those between-islands. We argue that volcanic eruptions and landslides after the merging of the palaeo-islands 3.5 Ma played key roles in generating genetic boundaries within Tenerife, with the palaeo-islands acting as refugia against extinction, and as cradles and sources of genetic diversity to other areas within the archipelago.

Keywords: ancestral areas, extinction, oceanic islands, palaeo-islands, volcanic refugia

Received 5 March 2015; revision received 16 June 2015; accepted 17 June 2015

Introduction

Due to their small size, discrete geographical boundaries, substantial environmental heterogeneity and buffered climates, oceanic islands represent excellent natural laboratories to examine the role of ecological adaptations vs. geographical isolation as drivers of diversification (Carson & Templeton 1984; Gillespie 2004). Patterns of genetic variation are expected to be hierarchical in islands, with between-island genetic differentiation stronger than within-islands (Bottin *et al.* 2005; García-Verdugo *et al.* 2010), because oceanic barriers are generally more effective than topographic barriers at promoting isolation in insular systems (Gillespie & Clague 2009). However, the birth and development of volcanic islands is usually followed by a large number of destructive events in the form of secondary eruptions, landslides, merging of palaeo-islands, etc. (Carracedo 1994; Fernández-Palacios *et al.* 2011). These events promote habitat fragmentation and the subsequent genetic isolation of populations, and they ultimately drive differentiation and speciation within-islands (Carson *et al.* 1990; Gillespie & Roderick 2002, 2014; Macías-Hernández *et al.* 2013). Therefore, volcanic archipelagos represent an ideal framework for studying patterns of diversification at dif-

Correspondence: Mario Mairal, Fax: +34914200157;
E-mail: mariomairal@gmail.com

[1]Equal contributions.

© 2015 John Wiley & Sons Ltd

ferent spatial scales (between- and within-islands) and also over different temporal levels (allopatric speciation, extirpations and recolonizations).

The Canary Islands are a volcanic archipelago formed by a chain of seven islands, located 110 km from the north-western coast of Africa. Their geological history and highly endemic biota have interested scientists since the early 19th century (von Humboldt 1814; Lyell 1855). The islands were formed in the last 21 Ma with an east to west pattern of chronological emergence due to a mantle plume (Carracedo *et al.* 1998; Zaczek *et al.* 2015) (Fig. 1a). They are separated by deep oceanic trenches and have never been connected to the mainland. The Canary Islands have long been considered refugia for continental lineages that have survived the climatic changes of the Late Cenozoic, but also as cradles of biodiversity where multiple in situ diversification events have taken place (Francisco-Ortega *et al.* 2000; Juan *et al.* 2000). In addition, these islands are regarded as a hotspot for plant diversity (Médail & Quézel 1997): approximately 40% of Canarian vascular plants are endemics (Santos-Guerra 2001) and the flora is generally characterized by high levels of interpopulation differentiation in comparison with other archipelagos (Francisco-Ortega *et al.* 2000; de Paz & Caujapé-Castells 2013). Recently, they have been proposed as reservoirs and sources of genetic diversity (García-Verdugo *et al.* 2015; Patiño *et al.* 2015).

Most studies on the Canarian flora have focused on the pattern of interisland diversification, particularly at the species level (Francisco-Ortega *et al.* 2002; Kim *et al.* 2008; Vitales *et al.* 2014a,b). Nevertheless, the complex topographies and long histories of avalanches and secondary eruptions of many of these islands are likely to have favoured within-island diversification (Juan *et al.* 2000; Brown *et al.* 2006). Tenerife has the most complex history of all the islands. It existed at first as three separate islands, dating back to the Late Miocene: Roque del Conde (11.9–8.9 Ma) in the southwest, Teno (6.2–5.6 Ma) in the northwest, and Anaga (4.9–3.9 Ma) in the northeast. Eruptive central volcanic episodes fused these islands 3.5 Ma and gave rise to the present island of Tenerife (Ancochea *et al.* 1990; Fig. 1b). The palaeo-islands remained thereafter relatively stable, whereas the central part of Tenerife continued to be active until 0.13 Ma (Ancochea *et al.* 1990, 1999; Cantagrel *et al.* 1999; Guillou *et al.* 2004; Carracedo 2014; see Fig. 1b). Interestingly, the three palaeo-islands of Tenerife, together with La Gomera—which has also remained geologically stable since the Pliocene—are presently home to the highest phylogenetic diversity and endemic richness of the Canarian Archipelago (Reyes-Betancort *et al.* 2008). These areas also harbour the best-preserved laurel forests, considered as an ancient, unique flora restricted to Macaronesia, and they share several endemic and restricted species (see Table S1, Supporting information). Besides their geological

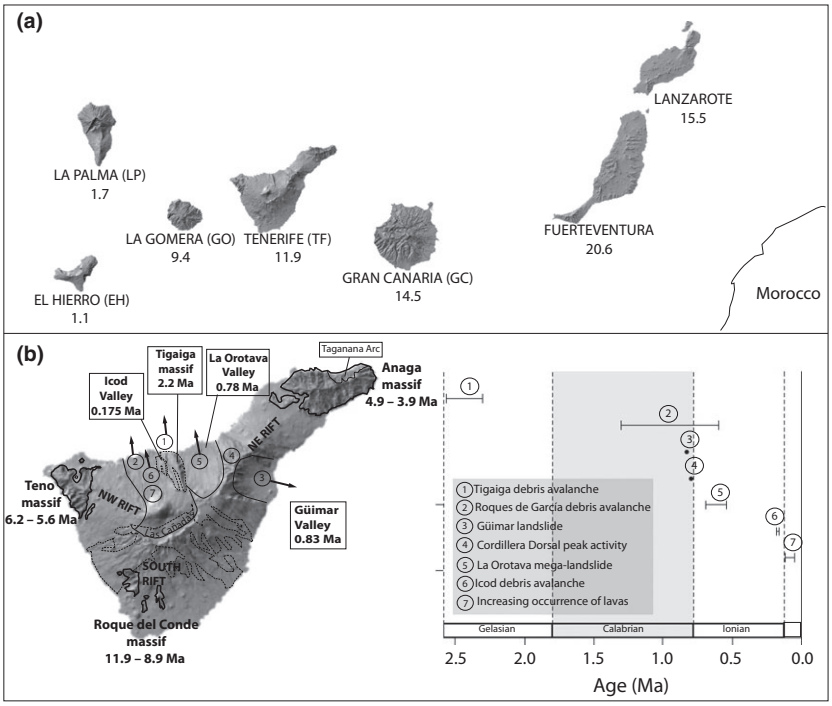


Fig. 1 (a) Geographical map of the Canary Island archipelago, with the code for each island used in the text and their age of emergence (million years, Ma). (b) Map of Tenerife showing main Pleistocene landslides and other geological events cited in the text. The old basaltic series corresponding to the palaeo-islands are shown in the three contoured massifs at the edges of the island. Lines show valleys created after landslides. Dashed lines indicate the post-erosional Las Cañadas volcano. Arrows indicate the directions of the landslides. The graph on the right shows the main Tenerife catastrophic events that are commented in the text; the numbers in the legend refer to those in the map. Ages and maps adapted from: Carracedo *et al.* 1998, 2007 ; Ancochea *et al.* (1999); Cantagrel *et al.* (1999); Guillou *et al.* (2004); Boulesteix *et al.* (2012); Zaczek *et al.* (2015).

© 2015 John Wiley & Sons Ltd

(volcanic) stability, the three palaeo-islands of Tenerife also exhibit a topographic complexity and variety of micro-climates that might have favoured their role as micro-refugia against climate- or human-induced extinction (Harter *et al.* 2015).

In reviewing the role of Tenerife palaeo-islands as refugia across several plant lineages, Trusty *et al.* (2005) found that species endemic to the palaeo-islands often occupied a derived position in the lineage's phylogeny. This position argued against the idea of these massifs as ancient refugia. However, other studies, especially in animals, have reported ages for divergence events between taxa endemic to these palaeo-islands, either at the species or at the intraspecies level (Juan *et al.* 1996, 2000; Dimitrov *et al.* 2008; Macías-Hernández *et al.* 2013; Puppo *et al.* 2014) that are contemporaneous or predate the age of merging of the precursor palaeo-islands 3.5 Ma (Ancochea *et al.* 1990). Few plant studies (Gómez *et al.* 2003; García-Verdugo *et al.* 2010) have focused on patterns of within-island genetic variation for widespread Canarian endemics, and none of them have provided estimates of lineage divergence times, which is necessary to relate within-island patterns to the island geological history. Species that are present in multiple islands ('multiple island endemics' or MIEs) are especially relevant to understand the role of palaeo-islands as undisturbed areas that have acted as reservoirs and sources of genetic diversity not only within- but between-islands.

Here, we study patterns of genetic diversity and the demographic and spatial history of a multiple island endemic, *Canarina canariensis* (L.) Vatke. This 'flagship' species of the Canary Islands, elected as its 'national flower' (Kunkel 1991), is a diploid (2n = 34) herbaceous plant that grows mostly in cleared areas surrounding the endemic laurisilva forest. It presently occurs in the central and western Canary Islands: Gran Canaria, Tenerife, La Gomera, La Palma and El Hierro. *Canarina canariensis* is a herbaceous plant that occasionally climbs on nearby plants; it is pollinated by generalist birds (Rodríguez-Rodríguez & Valido 2011) and its fleshy fruits are dispersed by vertebrates (Valido *et al.* 2003; Rodríguez *et al.* 2008). Genus *Canarina* belongs to tribe Platycodoneae, a basal group within family Campanulaceae (Mansion *et al.* 2012; Olesen *et al.* 2012; Wang *et al.* 2013; Mairal *et al.* 2015). In addition to *C. canariensis*, the genus comprises two other species inhabiting the Afromontane forests of Eastern Africa, *C. eminii* and *C. abyssinica*, being this an example of a wide, continental-scale disjunction of 7000 km spanning across the Sahara. Mairal *et al.* (2015) recently reconstructed the phylogeny and spatiotemporal evolution of *Canarina*. They inferred that *C. canariensis* diverged from its African relatives at the end of the Miocene (*c.* 7 Ma). This extraordinary temporal and spatial disjunction was explained as the result of vicariance and climate-driven extinction resulting in the fragmentation of an ancient widespread distribution. The colonization of the Canary Islands by the ancestors of *C. canariensis* apparently occurred much later, in the Pleistocene, probably from a now extinct and geographically closer North African population: the earliest event of population divergence is dated at only around 1 Ma (Late Pleistocene) and involved an east–west vicariance within Tenerife (Mairal *et al.* 2015).

This age and the presence of *C. canariensis* in several islands makes it an ideal candidate to evaluate patterns of within-island diversification in relation to the recent geological history of the archipelago. Our main aims were to: (i) determine the geographical distribution of genetic variation within *C. canariensis*, (ii) find evidence of extinction and diversification processes that may be related to geological events, (iii) find ancestral areas and reconstruct interisland migration events and (iv) examine the putative role of the palaeo-islands of Tenerife as refugia of genetic diversity, both relictual and recent.

Haplotype networks are commonly used in population-level studies because they provide a clearer picture of the reticulate relationships between genetic pools than a branching tree, especially when gene flow is present. These networks are often inferred using Statistical Parsimony (SP, Templeton *et al.* 1992) implemented in the software TCS (Clement *et al.* 2000), which allows estimation of the haplotype network while minimizing the number of mutation events differing among haplotypes. However, this method fails to incorporate the uncertainty associated with the network inference and therefore does not allow for statistical evaluation of alternative phylogeographical scenarios (Bloomquist *et al.* 2010). Moreover, unobserved events such as local population growth or past extinction of haplotypes may mislead inference in parsimony-based methods. Here, we compare results from TCS with those obtained from a model-based, Bayesian statistical method, Bayesian phylogeographic and ecological clustering (BPEC, Manolopoulou *et al.* 2011), which allows estimating the posterior probabilities for haplotype tree networks under a coalescent-based migration–mutation model (Manolopoulou & Emerson 2012). To our knowledge, this is the first study to use this method for island phylogeography.

Materials and methods

Population sampling and DNA extraction

Canarina canariensis has a significantly greater presence on Gran Canaria (GC) and Tenerife (TF), in comparison

© 2015 John Wiley & Sons Ltd

with La Gomera (LG), La Palma (LP) and El Hierro (EH). Seventeen populations of *C. canariensis* were sampled in several fieldtrips between 2009 and 2012: four in GC, eight in TF, two in LG, two in LP and one in EH. Where possible, we collected a minimum of 10 samples per population. To reduce inflation in gene descriptors due to biased sampling (Caujapé-Castells 2010), samples were collected from individuals scattered across the whole occupancy area of each population. DNA from 160 individuals and preserved in silica gel was extracted using the DNeasy Plant Mini Kit (QIAGEN Inc., Valencia, CA, USA), from 20 to 25 mg of silica-gel-dried leaves obtained from the fresh plant tissue collected from the field expeditions.

Chloroplast DNA sequencing

We selected three cpDNA intergenic spacers regions for sequencing; these markers have proven to be useful for intraspecific analyses of population structure (Mairal *et al*. 2015). We generated 432 new sequences: *rpl*32-*trn*L^{UAG} (144 sequences), *trn*S^{GCU}–*trn*G^{UCC} (144 sequences) and *pet*B^{1365}–*pet*D^{738} (144 sequences). PCR and sequencing protocols followed those of Mairal *et al*. (2015). The sources of the material examined, the GenBank Accession nos and full references are detailed in Table S2 (Supporting information).

Haplotype analyses

Sequences for each region were aligned using MAFFT 6.814b (Katoh *et al.* 2002), implemented in the software GENEIOUS PRO 5.4.4. (Biomatters Ltd., Auckland, New Zealand). Sequences were checked and manually adjusted where necessary by following alignment rules described in Kelchner (2000). We analysed the three sequenced regions as three data partitions to perform phylogenetic analyses. MRMODELTEST v.2.2 software (Nylander 2004) was used to determine the best fitting model of sequence evolution of each data partition.

Summary statistics for within-population genetic diversity were calculated as follows: the number of haplotypes H(n), haplotype diversity (Hd), nucleotide diversity π, nucleotide heterozygosity θ, and G_{ST} and the number of migrants per generation (Nm) were estimated for each population using DnaSP (version 5.10; Librado & Rozas 2009).

The relationships among lineages were investigated through haplotype network analysis, using 6–12 individuals from different populations and examining the three sequenced regions. Genealogical relationships among haplotypes were inferred via the statistical parsimony algorithm (Templeton *et al.* 1992) implemented in TCS 1.21 (Clement *et al.* 2000). The number of mutational steps resulting from single substitutions among haplotypes was calculated with 95% confidence limits, and gaps were represented as missing data.

The BPEC method (Manolopoulou *et al.* 2011; Manolopoulou & Emerson 2012) was implemented in the R package to identify genetically distinct geographical population clusters and ancestral locations. Like TCS, BPEC relies on parsimony in order to reduce the number of candidate trees to a manageable set. The BPEC method, unlike Standard Parsimony, fits a prior over all possible trees in order to identify trees with high posterior probability in a fully model-based framework, thus accommodating for uncertainty in haplotype relationships, which is one of the main criticisms of TCS (Knowles 2008). Each possible tree defines a set of possible migration events that may have led to the observed population substructure. Different scenarios of trees and migration events are explored through Markov chain Monte Carlo (MCMC), similar to the method proposed by Sanmartín *et al*. (2008) for estimating rates of interisland dispersal. Migration events were assumed to occur when a haplotype (with or without a mutation from its parent haplotype) migrates to a new geographical cluster. MCMC simultaneously estimates high probability trees, number of migration events and corresponding clusters. The method assumes that the migration rate and the population growth are constant. BPEC requires two main user-defined inputs: the maximum number of migrations (denoted as 'MaxMig' in the software) and the parsimony relaxation parameter used to reconstruct the set of possible trees (denoted as 'ds' in the software). MAXMIG allows the user to set the upper bound for the number of migration events and hence the maximum number of clusters (MaxMig + 1). Larger values include more models but require much greater computation time. As the ds value is increased, the parsimony assumption is relaxed: if two observed sets of sequences have an unobserved intermediate missing sequence (an unobserved mutation), then any pair of sequences of distance $\{1, \ldots, ds\}$ nucleotides will be considered as the 'missing path'. Two MCMC chains were run for 3 million iterations. The results were stable, with ds > 3 having no effect on the inferred tree, and similarly, any number of migrations above four converging to a 5–cluster model. The phylogeographical clustering obtained was superimposed upon a haplotype tree and used to estimate ancestral locations for migration events. As a further exploration, we divided the data set into two groups (eastern and western clades—for details, see *coalescent dating* section in results). Haplotypes sampled from Roque del Conde were quite divergent; thus, separate analyses were run in which we included and excluded this location from the two groups. We ran BPEC analysis on each of these four data sets

© 2015 John Wiley & Sons Ltd

(eastern and western groups with and without Roque del Conde).

Haplotype divergence times were estimated in BEAST v.1.7 (Drummond & Rambaut 2007). We carried out a first analysis under a strict clock model and a coalescent constant population tree prior, using a secondary age estimate (Mairal *et al.* 2015; normal prior: mean = 0.76 Ma, standard deviation (SD) = 0.327 Ma) to calibrate the root node of the '*C. canariensis* data set'; this included all haplotypes detected in our sampling ($N = 10$). This analysis gave us very large 95%HPD (high posterior density) credibility intervals and poor ESS for posterior age estimates, probably due to the low information content at the population level (see Results) and the presence of a single calibration point. We carried out a second analysis applying the 'nested dating approach' described in Mairal *et al.* (2015), in which a higher-level data set including representatives of all three species of *Canarina* and nine outgroup taxa was used to inform the clock rate of a linked population-level data set (*C. canariensis*) under a *mixed Yule-coalescent model* (Ho *et al.* 2005; Pokorny *et al.* 2011). The higher-level data set was calibrated with fossil-derived secondary age estimates (see Mairal *et al.* 2015), while the tree prior was unlinked to apply a coalescent constant size model to the population-level data set and a stochastic birth–death (Yule) prior to the species-level one (Mairal *et al.* 2015). The clock model was set to an uncorrelated log-normal prior to accommodate the change in mutation rate from species to populations, with a uniform distribution for the *ucld.mean* (10^{-4}–10^{-1}) and a default exponential distribution for the *ucld.stdev*; the substitution model was set to GTR+G; choice of priors was based on Bayes Factor comparisons using the path sampling method in BEAST (Baele *et al.* 2012); see Table S3 (Supporting information) for results from exploratory analyses to assess the reliability of our date estimates with reference to these settings. Two MCMC chains were run for 50 million generations, sampling parameters every 1000 generation. We used Tracer v1.6 (Rambaut *et al.* 2007) to verify the following: whether a stationary distribution was attained, whether there was convergence among chains and whether effective sample sizes (ESS values) were >200 for all parameters. A 10% burn-in of the sampled populations was discarded (5 million). Post-burn-in trees were summarized into a maximum clade credibility tree using TREEANNOTATOR v.1.6.1, with mean values and 95% credible intervals for nodal ages, and were visualized in FIGTREE 1.3.1 (Rambaut & Drummond 2009). The resulting age estimates from this second analysis exhibited considerably larger ESS values and narrower 95%HPD intervals than in the first and are the ones reported here.

Demographic history

Statistics used to describe demographic patterns may be biased by a strong genetic structure or lack of panmixia among populations. As we detected strong genetic structure in our data set (see results), we performed demographic analyses in subsetted data sets that were less genetically structured (including approximate panmictic populations). Three groups previously recognized by the haplotype network and BEAST analyses were used: ET-GC (including populations in east Tenerife and Gran Canaria), TENO-GO (including populations in Teno and La Gomera) and LP-EH (including populations in La Palma and El Hierro). For the same reasons as described above, the Roque del Conde population was included and discluded from the ET-GC and TENO-GO groups. Overall, we ran the analyses on five groups.

We used three different approaches to infer the demographic processes shaping the genetic structure of *C. canariensis*. First of all, to test for evidence of population expansion, we carried out a neutrality test—Fu and Li's tests (Fu & Li 1993; Fu 1996) and Tajima's *D* test (Tajima 1989)—for each population group. We used the DNAsp program, version 5.0 (Librado & Rozas 2009), and assessed the significance of each test with 10 000 coalescent simulations. Second, we plotted the mismatch distribution for each group using the observed number of differences between all pairs of sequences with the ARLEQUIN v.3.0 software (Excoffier *et al.* 2005). The goodness of fit of the observed mismatch distribution to the theoretical distribution under a constant population size model was tested with the raggedness index (*HRag*) (Harpending 1994). Third, (i) we created the extended Bayesian skyline plot (EBSP), implemented in BEAST, for each population group, and (ii) we performed EBSP analyses for each of the 16 population groups. For each group from (i) and (ii): two independent chains were run simultaneously for 150 million generations, sampling every 1000 generations; a strict clock model was used, whereas all other parameters were set identical to those described above for the nested dating analysis; the root node was calibrated using a normal prior with a mean age estimate and 95% high posterior density (HPD) credible intervals obtained from this analysis.

AFLP fingerprinting

For the AFLP analysis, we used a total of 97 individuals from 10 populations, which covered all of the islands: one population from Gran Canaria, five from Tenerife, one from La Gomera, two from La Palma and one from El Hierro. Laboratory molecular protocols for the AFLP

© 2015 John Wiley & Sons Ltd

analysis (Vos *et al.* 1995) were implemented using the AFLP plant mapping kit (Applied Biosystems®, Foster City, CA, USA). To select the appropriate primers, we first carried out a pilot study combining fluorophores and restriction enzymes for five geographically distant individuals (one per island), using in total 32 primer combinations. One sample from each individual was duplicated as a blind sample to test for reproducibility and contamination. Reproducibility and the number of alleles per sample were calculated by choosing three combinations of primers: 1-*Eco*RI$_{6\text{-FAM}}$-ACT/*Mse*I-CAA, 2-*Eco*RI$_{VIC}$-AGG/*Mse*I-CTA and 3-*Eco*RI$_{VIC}$-AGG/*Mse*I-CTT and using the GENEMAPPER v3.7 software (Applied Biosystems). These three primer pairs showed high reproducibility and homogeneously scattered bands and produced polymorphic AFLP profiles and clear fragments. In the digestion phase, samples of DNA were digested with the restriction enzymes EcoRI and MseI and linked to the primers EcoRI 5′-CTCGTAGAC TGCGTACC-3′/5′AATTGGTACGCAGTCTAC-3′) and MseI (5′- GACGATGAGTCCTGAC-3′/5′-ATCTCAGGA CTCAT-3′). The three different AFLP reactions were as follows: (i) restriction and ligation in a single reaction; (ii) and (iii) consecutive PCR amplifications (preselective and selective). PCR products were checked on 1% agarose gels.

AFLP data analyses

The resulting AFLP fragments were analysed using the GENEMAPPER 3.7 software. Peaks were recorded in 100–500 base pairs ranges. Shorter fragments were discarded because the majority of this size class have a high chance of being nonhomologous fragments (Vekemans *et al.* 2002). For each primer combination, an automated size detection and peak binning was employed followed by manual editing of bins to exclude shoulder peaks and unreliable loci (variation between replicates). Peak height data were then exported and loaded into the R package AFLPSCORE version 1.4a (Whitlock *et al.* 2008), and the AFLP profiles were scored and the error rates were estimated. These rates were below the critical bound of 5% indicated in previous reports (Bonin *et al.* 2004) for each primer combination. Data reliability was assessed through comparison of duplicates, from one or two individuals per population. Data were converted into binary presence/absence scores for each locus. The resulting AFLP presence/absence matrix was analysed using a selection of different analyses. The AFLPSURV v.1.0 software (Vekemans *et al.* 2002) was used to estimate demographic statistics such as Nei's gene diversity (Hj), pairwise differentiation among subpopulations (F_{ST}) and the percentage of polymorphic fragments per population (P) (Nei & Li 1979; Lynch & Milligan 1994). This was done under the assumptions of either the Hardy–Weinberg equilibrium or partial self-fertilization, based on a previous study on the reproductive biology of *C. canariensis* (Rodríguez-Rodríguez & Valido 2011). A Bayesian method in AFLPSURV was also used to estimate allelic frequencies through employment of a nonuniform prior distribution (Zhivotovsky 1999). Ten thousand permutations were run to calculate the F_{ST} parameter from which genetic distances between individuals, populations and geographical groups were calculated. To locate genetic clustering of individuals within the AFLP data set, a pairwise similarity matrix for all individuals using the Dice's coefficient as similarity distance was constructed, and the resulting matrix subjected to a principal coordinates analysis (PCA) implemented in NTSYS v.2.1 (Rohlf 1998). Next, genetic relationships among samples were visualized in SPLITSTREE v.4.10 (Huson & Bryant 2006) using neighbour-net analysis through the use of the split decomposition method. Finally, to quantify the amount of genetic differentiation attributable to geographical and population subdivision, a hierarchical analysis of molecular variance was performed using ARLEQUIN v.3.0 software (Excoffier *et al.* 2005). Exploratory analyses were performed considering, alternatively, islands and palaeo-islands as geographical units in order to investigate the distribution of genetic variance attributable to oceanic barriers.

Bayesian clustering methods implemented in STRUCTURE v.2.3 (Pritchard *et al.* 2000; Falush *et al.* 2007) were used to assess the genetic structure of populations. This model-based approach assumes that loci are in Hardy–Weinberg equilibrium and linkage equilibrium within populations. Analyses were performed under admixture conditions and correlated allele frequencies between groups. 500 000 MCMC generations (plus a burn-in of 100 000) were run for K values of 1–10, with 10 repetitions for each. The most likely K value was determined by the method from Evanno *et al.* (2005), which is implemented in STRUCTURE HARVESTER (Earl 2012). We explored other values of K to detect further genetic substructure of populations, especially in Tenerife. To test the effect of the spatial distance on the genetic structure of the populations, correlations between genetic (measured as F_{ST}) and spatial distances between pairs of populations were determined using the Mantel permutation procedure implemented in NTSYS v. 2.1. The genetic distance matrix used was based on the presence/absence matrix; the geographical distance matrix was based on the absolute distances between the geographical coordinates for each collected population. In addition, to identify possible geographical locations acting as major genetic barriers among *C. canariensis* populations, we computed barriers on a Delaunay triangulation using Monmonier's algorithm in BARRIER

© 2015 John Wiley & Sons Ltd

v.2.2 (Manni *et al.* 2004). The significance was examined by the mean of 1000 bootstrapped distance matrices obtained using AFLPSURV. Only barriers with support >96% were considered as significant.

Results

Haplotype network analysis and coalescent dating

The *pet* B^{1365}–*pet* D^{738} region consisted of 937 sites, *rpl*32–*trn*L^{UAG} of 654 sites and *trn*S^{GCU} –*trn*G^{UCC} of 658 sites. The final concatenated data matrix consisted of 144 sequences and 2249 nucleotide sites, of which nine were polymorphic (Table S4, Supporting information). Among the 144 individuals sampled from the 16 populations, we observed 10 different plastid DNA haplotypes (H1 to H10, Fig. 2a) with haplotype diversity of 0.6632. There were three haplotypes that dominated and were clearly geographically delimited: H1, H9 and H3. The most frequently observed haplotype was H1 (52.08%), which was dominant in GC and east Tenerife (ET). Haplotype H9 was restricted to West Tenerife (WT) and LG. Haplotype H3 was present in the western islands of LP and EH. Populations with the commonest and rarest haplotypes (frequencies ≤ 2%) were in the most rugged areas of Tenerife: the palaeo-islands of Anaga (An-W, An-C, An-E—exclusive haplotypes H2 and H5), Teno (Ten-A, Ten-Ep—exclusive allele H8 and allele H10 shared with LG) and Roque del Conde (Conde—exclusive H7). The small western islands presented one unique haplotype, each: LP (LP-Lt—H4) and EH (H6).

Summary statistics for within-population genetic diversity are shown in Table 1. Populations situated in the three palaeo-islands [Anaga (An-W), Teno (Ten-A) and Roque del Conde (Conde)] showed the highest values for the number of haplotypes H(n), haplotype diversity H(d), nucleotide diversity π and nucleotide heterozygosity θ. The least diverse population areas were found in Gran Canaria and outside the palaeo-islands regions of Tenerife. G_{ST} and N_m values indicated high genetic cohesion between some islands: east Tenerife populations were highly connected with Gran Canaria populations; west Tenerife with La Gomera; and La Palma with El Hierro (Table S5, Supporting information). Interestingly, the highest G_{ST} values and lowest Nm were found between populations situated east and west of Tenerife, with the exception of Anaga An-W. This is the only population in east Tenerife that showed some genetic cohesion with populations in west Tenerife, in particular with the Teno and Adeje massifs.

The haplotype network constructed with TCS (Fig. 2b) shows a loop involving three dominant haplotypes, with haplotype H1 as the ancestral root haplotype according to coalescent criteria on haplotype frequency (Templeton *et al.* 1992). Dominant haplotypes at each side of Tenerife (East H1 and West H9) were separated by five mutational steps, with haplotype H10 as intermediate. Haplotype H3, present in the western islands, was separated by four mutational steps from H1 and three from H9.

BPEC results are summarized in Table 2. With mutational step limit equal to 10 and the maximum number of migrations equal to 8, we allowed for high gene flow, genetic divergence and numerous hidden mutations, which separate the haplotype clusters. The clustering with high posterior probabilities (pp = 1) showed the existence of five clusters (contour regions in Fig. 3a). The total number of haplotypes was 10 plus one missing haplotype (Fig. 3b). Location of the ancestral populations for migration events and inference of the root node were consistent across exploratory analyses of the

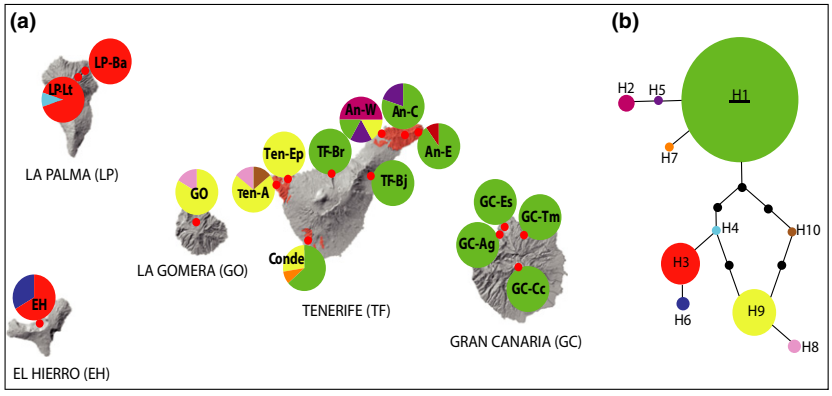


Fig. 2 Haplotype distribution and network inferred for the chloroplast markers by TCS. (a) Pie charts show the geographical location of populations and the frequency of occurrence of each haplotype, and circle size is proportional to population size. Population codes are given in Table S2 (Supporting information). (b) Statistical Parsimony network inferred by TCS. Each haplotype is shown in a different colour; circle size is proportional to its frequency among populations.

© 2015 John Wiley & Sons Ltd

Table 1 Descriptors of within-population genetic diversity in the cpDNA haplotypes and AFLP markers for each population studied of *C. canariensis*

	Haplotypes							AFLPs		
Island (Palaeo-island)	Population	No samples	Haplotype	H (*n*)	H (d)	π	θ	No samples	No of polymorphic fragments (% in brackets)	Hj (SE)	Number of private fragments
Gran Canaria	GC-Cc	8	H1	1	0	0	0	8	409 (68.7)	0.2344 (0.0085)	0
Gran Canaria	GC-Tm	11	H1	1	0	0	0	NA	NA	NA	NA
Gran Canaria	GC-Es	10	H1	1	0	0	0	NA	NA	NA	NA
Gran Canaria	GC-Ag	11	H1	1	0	0	0	NA	NA	NA	NA
Tenerife (Anaga)	An-W	6	H1, H2, H5, H9	4	0.8	0.00112	0.00136	7	401 (67.4)	0.2677 (0.0088)	4
Tenerife (Anaga)	An-E	10	H1, H2	2	0.2	0.00018	0.00031	10	345 (58.0)	0.2303 (0.0080)	0
Tenerife (Anaga)	An-C	11	H1, H5	2	0.327	0.00015	0.00015	11	375 (63.0)	0.2453 (0.0078)	1
Tenerife	TF-Bj	4	H1	1	0	0	0	NA	NA	NA	NA
Tenerife	TF-Br	6	H1	1	0	0	0	NA	NA	NA	NA
Tenerife (Roque del Conde)	Conde	11	H1, H9, H7	3	0.618	0.00121	0.00091	NA	NA	NA	NA
Tenerife (Teno)	Ten-Ep	8	H9	1	0	0	0	9	351 (59.0)	0.2449 (0.0082)	0
Tenerife (Teno)	Ten-A	7	H8, H9, H10	3	0.524	0.00038	0.00054	7	415 (69.7)	0.2997 (0.0088)	2
Gomera	GO	12	H8, H9	2	0.303	0.00013	0.00015	12	473 (79.5)	0.3012 (0.0077)	1
La Palma	LP-Ba	10	H3	1	0	0	0	9	285 (47.9)	0.1923 (0.0080)	0
La Palma	LP-Lt	10	H3, H4	2	0.2	0.00009	0.00016	10	348 (58.5)	0.2309 (0.0079)	0
El Hierro	EH	9	H3, H6	2	0.5	0.00022	0.00016	9	303 (50.9)	0.2115 (0.0082)	1

H(*n*), number of haplotypes; H(d), haplotype diversity; π, nucleotide diversity; θ, nucleotide heterozygosity; Hj (se), Nei's gene diversity (standard error). Geographical locations for population codes are shown in Fig. 2. The most diverse population in each palaeo-island is shown in bold.

© 2015 John Wiley & Sons Ltd

complete data set. The root node was inferred as missing (H11—Fig. 3b) and carried a high uncertainty. Haplotypes H1, H2 and H8 also carried significant posterior mass probabilities (Table 2). The most likely source for ancestral migration events within- and between-islands were populations located on the following: Conde (pp = 0.13), An-W (pp = 0.10) and LP-Lt (pp = 0.086). The subsetting analysis, including Conde, always recovered this population as the ancestral area with the highest probability. Without Conde, other palaeo-island populations were recovered as ancestral areas: Ten-A for the Western clade; and the three populations of Anaga for the Eastern clade, with An-W the most probable (Table 2; Fig. S1, Supporting information).

The Bayesian chronogram of haplotypes showed a geographical pattern of divergence congruent with the groups detected by TCS and BPEC (Fig. 4). The crown age or first divergence event among *Canarina* haplotypes was estimated to be 0.878 Ma (95% HPD = 0.452–1.365), corresponding to the Mid-Pleistocene period; it divided haplotypes into an eastern and western clade. The first divergence event in the eastern clade was estimated at 0.255 Ma (95% HPD = 0.041–0.633), while that in the western clade was dated older: 0.522 Ma (95% HPD = 0.161–0.967), separating La Palma and El Hierro from Teno and La Gomera. Mean age values and 95% HPD intervals for the BEAST analysis are shown in Table S6 (Supporting information).

Demographic analyses

Negative values of Fu's and Tajima's indicated a recent population expansion, although these values were only significant for the Fu's test of the East group (Table S7, Supporting information). The frequencies of pairwise

Table 2 Summary of results from the Bayesian phylogeographic and ecological clustering (BPEC) analyses. Selected values for parameters MaxMig (maximum number of migrations) and ds (relaxed parsimony assumption) are shown. Ancestral locations (Anc. Loc.) for migration events are shown with their associated posterior probabilities

	Maxmig	ds	Anc. Loc. 1	Anc. Loc. 2	Anc. Loc. 3	Root node	Root node probability	No of clusters
Total	8	10	Conde (0.13)	An-W (0.10)	LP-Lt (0.086)	H11 (missing)	0.17	5
Western Clade (WT + LP + GO + EH)	4	3	Ten-A (0.40)	Lp-Lt (0.31)	EH (0.18)	H11 (missing)	0.18	3
Eastern Clade (ET + GC)	1	3	An-W (0.212)	An-C (0.122)	An-E (0.116)	H11 (missing)	0.16	2
Western Clade + Conde	4	3	Conde (0.416)	Ten-A (0.197)	GO (0.190)	H11 (missing)	0.18	3
Eastern Clade + Conde	1	3	Conde (0.23)	An-W (0.17)	An-C (0.086)	H11 (missing)	0.16	3

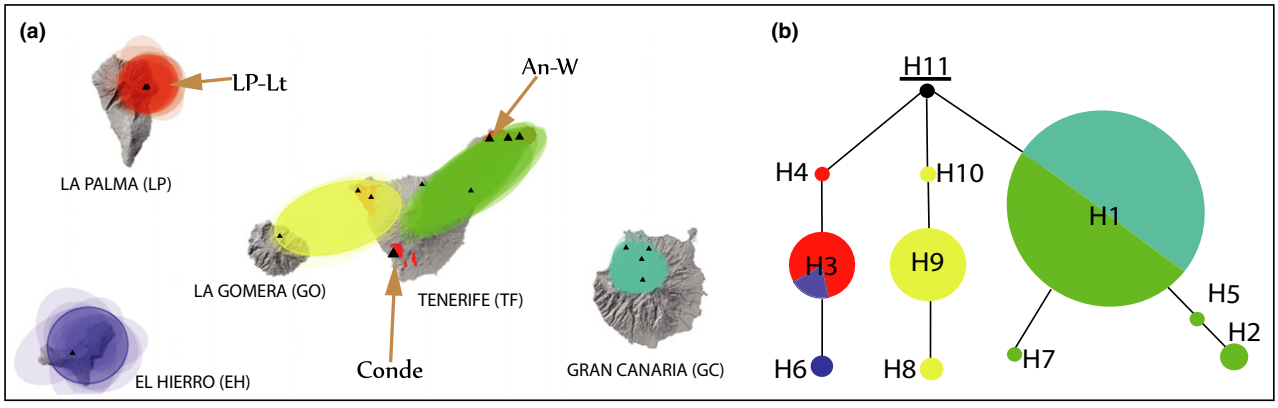


Fig. 3 Results from the Bayesian phylogeographic and ecological clustering (BPEC) analyses for the cpDNA markers. (a) Phylogeographical clusters (coloured blobs) and ancestral location for migration events (denoted by arrows). The contour regions are centred at the 'centre' of each population cluster, and the shaded areas show the radius of 50% concentration contours around it. Locations situated beyond the clusters could also belong to these clusters, but with low probability; in the case of Conde, it suggests a mixed composition, with this population as source of migrant alleles to the east and western clusters. (b) Haplotype network receiving the highest posterior probability. The small black circle H11 indicates an unobserved (missing) ancestral haplotype.

© 2015 John Wiley & Sons Ltd

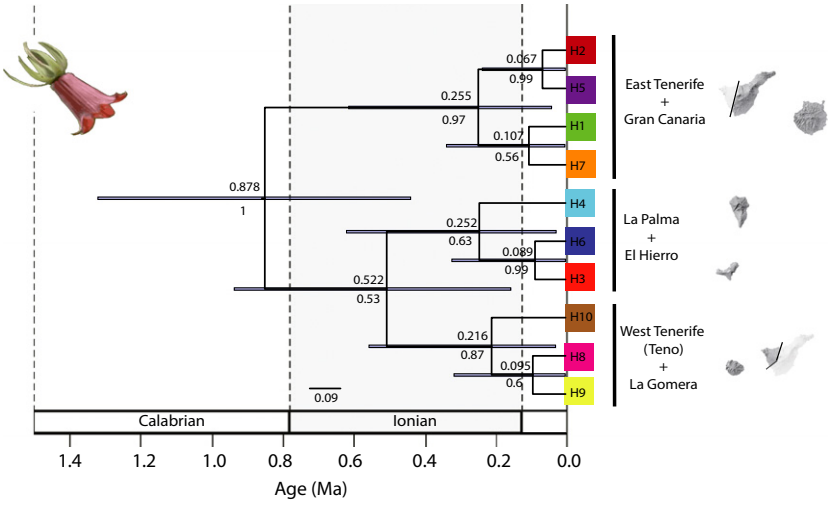


Fig. 4 Maximum clade credibility (MCC) tree obtained from the BEAST analysis of cpDNA haplotypes, showing mean ages (above branches) and 95% HPD credible intervals. Numbers below branches indicate Bayesian posterior clade support values. Codes for tips (H1 to H10) correspond to the haplotypes shown in Fig. 2.

differences in the mismatch distribution analysis resulted in unimodal distributions that were consistent with an expansion model. When we include the population of Conde in the East group, we obtained a second small peak. The raggedness statistics derived from the mismatch distribution were not significant and thus failed to reject the null hypothesis of recent population expansion. EBSP indicated a constant population size (Fig. S2, Supporting information). EBSP for each of the individual populations also obtained a constant population size (results not shown).

AFLP polymorphism, genetic diversity and structure

These results are summarized in Table 1. The final data set, after scoring, comprised 572 loci from 97 individuals and 10 populations. For AFLP analyses, we removed populations from Gran Canaria and Roque del Conde, which had been extracted with a different (less rigorous) method (CTAB) to standardize the quality of extracted DNA. Hj and P were higher in the La Gomera (GO) population, followed by the Teno (Ten-A) population in west Tenerife. The largest number of private fragments was detected in Tenerife (seven fragments). In La Gomera and El Hierro islands, a single fragment was detected. In Tenerife, five private fragments were detected in the Anaga massif (specifically in the An-W population—four fragments) and two in the Teno massif (Table 1). F_{ST} values with the nuclear data (Table S8, Supporting information) were consistent with the results obtained from the chloroplast markers (see above G_{ST} and Nm) and showed the same genetic cohesions between-islands and palaeo-islands.

A PCA differentiated three groups (Fig. S4, Supporting information): (i) east Tenerife populations; (ii) west Tenerife populations; and (iii) the An-W population, which demonstrate an intermediate position between (i) and (ii). The split network analysis outlined the same divergences between the two well separated groups, including the West group An-W population (Fig. 5b).

Hierarchical AMOVA analyses showed the largest proportion of genetic variation to be found among groups 4 and 5 (Table 3; Table S9, Supporting information). By analysing each island separately, we observed that differentiation within-islands (24.27%) is greater than between-islands (3.62%). A further analysis considering islands and palaeo-islands as separate units showed that differentiation within-islands and within-palaeo-islands is lower (11.30%) than between-islands and between-palaeo-islands (16.63%). According to the method of Evanno *et al.* (2005), STRUCTURE indicated that the most likely number of genetic clusters $K = 2$ ($\Delta K = 460$) represented the optimal number of Bayesian groups within *C. canariensis*, separating the east populations from the west, but also detected the presence of admixture in the intermediate populations. K levels $K = 4$ and $K = 5$ (the latter identified as $K = 6$, with 5 defined clusters plus one 'ghost' cluster with no individuals assigned so it was ignored, see Guillot *et al.* 2005) revealed a more complex genetic structure in these intermediate populations, resulting in additional clusters (Fig. 5a; Fig. S5, Supporting information). The STRUCTURE analysis using only Tenerife populations also delimited two groups (inset $K = 2$ in Fig. 5a; Fig. S5b, Supporting information). Overall, the results revealed a strong interisland genetic structure (Gran Canaria – Anaga massif and La Palma—El Hierro), admixture phenomena (among An-W, Teno populations and La Gomera) and an even more complex substructure within the island of Tenerife. The STRUCTURE sublevels ($K = 4$ and $K = 5$) were consistent with the AMOVA analyses. Only La Gomera differed between analyses; this is explained due to its high genetic variability and mixed composition. These groups are also congruent

© 2015 John Wiley & Sons Ltd

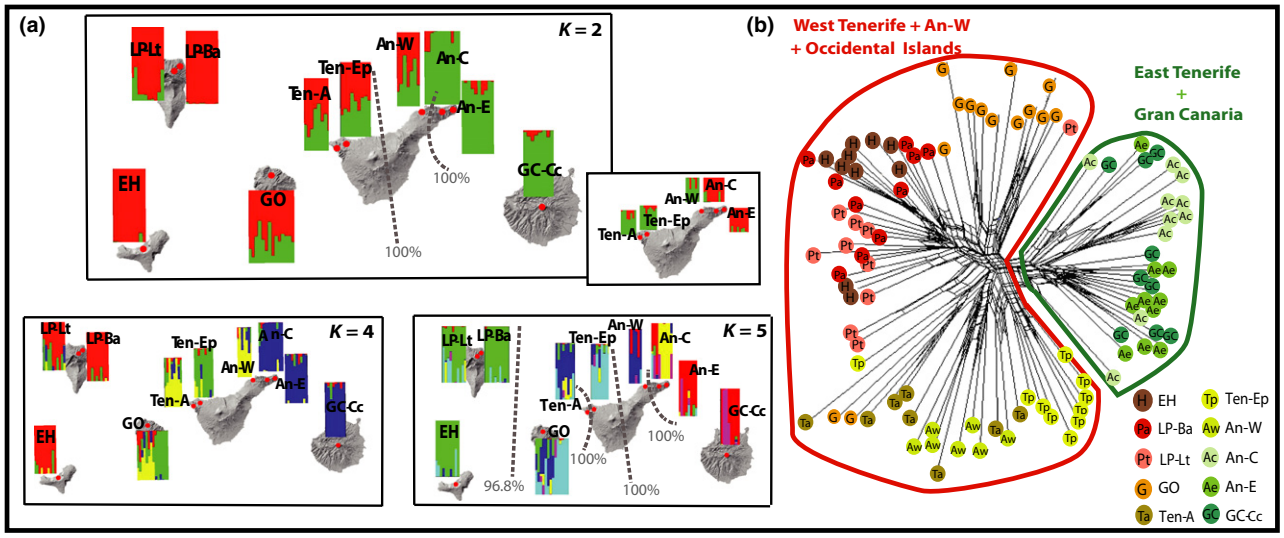


Fig. 5 Results from the analysis of AFLP markers. (a) Histograms showing the Bayesian clustering of individuals within populations (STRUCTURE); colours represent the proportion of individual membership to each inferred Bayesian group. Dotted lines indicate barriers to gene flow and their percentage, as inferred by BARRIER. (b) Split network with points coloured according to location, as indicated in the legend. Codes for populations found in Table S2 (Supporting information).

with the BARRIER results, which revealed two or four major boundaries (Fig. 5a for $K = 2$ and $K = 5$). All barriers showed values of 100% except LP-EH (96.8%). No linear relationship was found between pairwise F_{ST} and geographical distance with the Mantel analysis ($r = 0.42$, $P = 1$).

Discussion

Palaeo-islands as refugia against volcanic (catastrophic) extinctions

The strong genetic structure often detected in island endemics has been postulated to respond to several interconnected physical and biotic factors (Stuessy *et al.* 2014); foremost among them are geographical isolation and extinction-recolonization processes (Carson *et al.* 1990; Gübitz *et al.* 2005; Macías-Hernández *et al.* 2013). The strongest genetic structure and levels of genetic variation in *C. canariensis* were detected between two of the ancient palaeo-islands that currently form Tenerife: Teno in the west and Anaga in the east. Populations in other islands were genetically associated with these two lineages: La Palma, La Gomera and El Hierro to Teno; and Gran Canaria to Anaga. The fact that these patterns are shared between chloroplast (cpDNA, Fig. 2) and nuclear (AFLPs, Fig. 5) markers supports a long history of isolation among populations (Zink & Barrowclough 2008). Furthermore, the hierarchical AMOVA analysis based on the AFLP data showed that genetic variation was notably higher among populations within Tenerife than those located in different islands (Table 3). This pattern has been found in a few other Canarian endemics (*Olea europaea guanchica* and *Pinus canariensis*), but

Table 3 Hierarchical analysis of molecular variance (AMOVA) for *C. canariensis* based on allelic variation at different levels: (A) among groups, (B) among populations within groups and (C) within populations

		Levels			*F*-statistics		
AMOVA groups	No of groups (*K*)	A	B	C	Fsc	Fst	Fct
[GC + Anaga] [An-W + Ten-A] [Ten-Ep] [GO + LP + EH]	4	17.44	13.06	69.5	0.15820	0.305	0.17439
[GC+ An-E] [An-C] [Ten-Ep] [An-W + Ten-A] [GO + LP + EH]	5	18.92	11.21	69.88	0.13819	0.30123	0.18918
[GC] [TF] [GO] [LP] [EH]	5	3.62	24.27	72.10	0.25186	0.27897	0.03624
[GC] [An-E + An- C] [An-W] [Ten-Ep] [Ten-A] [GO] [LP] [EH]	8	16.63	11.30	72.07	0.13558	0.27934	0.16631

Geographical locations of populations are shown in Fig. 2.

© 2015 John Wiley & Sons Ltd

usually not as marked as here (<10%: Gómez *et al.* 2003; García-Verdugo *et al.* 2010; see review in García-Verdugo *et al.* 2014). Interestingly, considering the palaeo-islands as separate areas reversed the pattern (Table 3). Together with our BARRIER analyses (Fig. 5a in $K = 5$), these results support the hypothesis that the geographical distribution of genetic diversity in *C. canariensis* is structured around the palaeo-islands and that these ancient massifs could be considered as separate units in phylogeographical analyses, a sort of 'islands within-islands'. It indicates that geological barriers within-islands—for example, those associated to catastrophic volcanic events—have probably been more important in structuring patterns of genetic differentiation within species than the oceanic barriers separating the islands (García-Verdugo *et al.* 2014).

Unfortunately, we could not obtain AFLP data for the population in Roque del Conde. Given the mixed composition of the chloroplast compartment in this population (Fig. 2), it is possible that inclusion of this population in our analysis would have led to higher levels of genetic admixture and lower K values in the STRUCTURE analyses—though given the marked east/west split among the remaining populations (Fig. 5), this is unlikely. On the other hand, the two most frequent haplotypes, H1 and H9, in the Conde population are also the most frequent or dominant within the eastern and western clades, respectively, while the divergence of these two haplotypes in the BEAST tree (Fig. 4) can be traced back to the basalmost split in *C. canariensis* (0.8 Ma). This, together with the presence of a unique haplotype (H7) and the fact that the Conde population is identified as the ancestral source of westward and eastward migration events in the BPEC analyses, suggests that the admixture detected in the chloroplast compartment for Roque del Conde is of ancient rather than recent origin. It is thus possible that including this population in the AFLP analysis would have increased rather than decreased levels of genetic diversity within Tenerife for *C. canariensis*.

Which might have caused this level of within-island genetic structure? Many Canarian plant and animal taxa include sister lineages endemic to the Tenerife palaeo-islands (Juan *et al.* 2000; Báez *et al.* 2001), especially among laurel forest species from Teno and Anaga (Table 4). While in some cases, the divergence between these taxa is found at the species level (e.g. *Trechus*, Contreras-Díaz *et al.* 2007; *Pericallis*, Jones *et al.* 2014) and/or predates the merging of the palaeo-islands (*Micromeria*, Puppo *et al.* 2014; *Pholcus*, Dimitrov *et al.* 2008); in others, it is observed within species (e.g. *Tarentola delalandii*, Gübitz *et al.* 2000) and/or postdates the merging of the palaeo-islands (*Eutrichopus*, Moya *et al.* 2004; *Calathus abaxoides*, Emerson *et al.* 1999). The fact that the pattern of divergence is not contemporary across taxa suggests that the role of Tenerife palaeo-islands in structuring genetic variation has been long lasting. In *C. canariensis*, the basal divergence between the basal eastern and western lineages was dated as 0.878 Ma (Fig. 4) substantially, postdating the age of the merging of the palaeo-islands (*c.* 3.5 Ma, Ancochea *et al.* 1990). This 'young' east–west divergence might be explained by historical dispersal events between forest patches followed by in situ diversification. Alternatively, it could be explained by allopatric speciation (vicariance) driven by catastrophic volcanic events within a previously more widespread distribution with low or no reconnection. This might explain the 'young' (<3.5 Ma) east–west Tenerife divergence found in *Canarina*, and seen in other taxa (Table 4). The reciprocal monophyly, similar haplotype diversity levels and deep temporal divergence found between the eastern and western clusters of *C. canariensis* in Tenerife (Fig. 4) favour the vicariant, allopatric scenario. In the last 1 million years, several major landslides and volcanic events have affected the north of Tenerife, extending from the summit to the coast (Boulesteix *et al.* 2012, 2013; Carracedo 2014), for example the Güimar and La Orotava mega-landslides (Fig. 1b). These events could have fragmented the ancient laurel forest corridor that extended across the northern flank of the island (Moya *et al.* 2004), interrupting gene flow within species associated with this laurel forest (e.g. *C. canariensis*) and promoting differentiation among populations. Several studies have supported debris avalanches as important factors driving diversification within terrestrial Canarian organisms (Juan *et al.* 2000; Brown *et al.* 2006; Macías-Hernández *et al.* 2013); others have reported a temporal divergence within northern Tenerife (Thorpe *et al.* 1996; Moya *et al.* 2004) that is similar to the one found in *C. canariensis* and corresponds well with the time of the La Orotava mega-landslide (Ancochea *et al.* 1990; Boulesteix *et al.* 2013).

Whereas central Tenerife was almost completely covered by catastrophic events until as recently as 0.13 Ma (Ancochea *et al.* 1999), the three palaeo-islands of Teno, Anaga and Roque del Conde remained geologically stable since the mid-Pliocene (see Fig. 1). This suggests that these areas could have acted as refugia, allowing the survival of populations that otherwise disappeared in other parts of the island that were affected by volcanic activity. Support for this suggestion comes from the population genetic analysis. According to the central–marginal hypothesis (Eckert *et al.* 2008), spatial structure and genetic diversity should be higher in areas that have acted as refugia for the preservation of genotypes that went extinct in other areas and generally for the long-term persistence of populations (Hewitt 2000;

© 2015 John Wiley & Sons Ltd

Table 4 Phylogeographical breaks with divergence times reported in the literature between the palaeo-islands of Tenerife

Organism	Phylogeographical disjunction and groups	Divergence (Ma)	Markers	References
Species complex in darking beetle *Pimelia*	East Tenerife/West Tenerife	5.5 Ma	Mitochondrial (COI) and nuclear (ITS-1)	Juan *et al.* (1996)
Lizard *Gallotia galloti*	Western/North-eastern lineages	0.7 Ma	Mitochondrial cytochrome b	Thorpe *et al.* (1996)
Beetle *Calathus abaxoides*	Teno/Anaga	350 000 years	Two mitochondrial (COI and COII)	Emerson *et al.* (1999)
Skink *Chalcides viridanus*	Teno/Anaga	1.1 Ma	Two mitochondrial (12S and 16S)	Brown *et al.* (2000)
Gecko *Tarentola delalandii*	Clade 1. Teno + Roque del Conde/Anaga. Clade 2. Teno/Roque del Conde	Clade 1. 9–10 Ma Clade 2. 7.6 Ma	Mitochondrial cytochrome b	Gübitz *et al.* (2000)
Mite *Steganacarus carlosi*	Clade 1. Teno/Anaga. Clade 2. Roque del Conde/Anaga	Clade 1. 3.2 Ma Clade 2. 25–3.6 Ma	Mitochondrial cytochrome oxidase 1	Salomone *et al.* (2002)
Beetles *Eutrichopus*	Teno (*E. gonzalezi*)/ Anaga (*E. canariensis*)	0.7 Ma	Mitochondrial (COII marker)	Moya *et al.* (2004)
Species complex in beetle *Tarphius*	Teno/Anaga	1.2 Ma (1–1.4)	Two mitochondrial (COI and COII)	Emerson & Oromí (2005)
Beetle *Trechus*	Clade 1. Anaga + Teno (*T. antonii*, *T. tenoensis*, *T. felix*)/Anaga (*T. fortunatus*) Clade 2. *T. flavocintus*; Teno/East Tenerife	Clade 1. 1.73 Ma (HPD: 1.48–2.01) Clade 2. approx 0.75 Ma	Part of mitochondrial genes cytochrome oxidase I and II (Cox1 and Cox2),and nuclear (ITS 2)	Contreras-Díaz *et al.* (2007)
Spider *Pholcus*	Anaga (*P. malpaisensis*, *P. knoeseli*)/Teno (*P. intricatus*, *P. mascaensis*/*P. tenerifensis* and *P. roquensis*)	3.93 Ma (HPD: 2.2–5.88)	Four mitochondrial (CO1, 16S, NADH and tRNAleu). Morphological data	Dimitrov *et al.* (2008)
Grasshopper *Arminda brunneri*	Anaga + Güimar/Teno + Roque del Conde	1–0.17 Ma	Two mitochondrial (12s rRNA, ND5) and two nuclear gene fragments (28s rRNA, ITS2)	Hochkirch & Goerzig (2009)
Spider *Dysdera verneaui*	Teno/Anaga	3.94 Ma (HPD: 5.1–2.7)	Mitochondrial (*cox*1)	Macías-Hernández *et al.* (2013)
Plant *Pericallis*	Teno (*P. echinata*)/ Anaga (*P. tussilaginis*)	2.87 Ma (HPD: 1.55–4.76)	Nuclear ITS	Jones *et al.* (2014)
Plant *Micromeria*	Teno (*M. densiflora*)/Anaga (*M. teneriffae*, *M. glomerata* and *M. rivas-martinezii*) + Central group	5.2 Ma	8 nuclear loci	Puppo *et al.* (2014)

Tzedakis *et al.* 2013; Feliner 2014; Gavin *et al.* 2014). The higher number of ancestral and endemic alleles, private fragments and larger heterozygosity levels exhibited by the populations of the palaeo-islands of Tenerife (Table 1) are congruent with the idea that these massifs acted as reservoirs of ancient genetic diversity and as refugia against volcanically induced extinction. Interestingly, La Gomera, an island that has been geologically quiescent since the Pliocene (Carracedo & Day 2002), shows the highest Hj and percentage of polymorphic nuclear DNA fragments (Table 1), suggesting that this island might have acted in a similar way to the palaeo-islands of Tenerife.

Further support for the extinction hypothesis comes from the BPEC analysis. Theoretical predictions of coalescent theory states that high-frequency haplotypes have been present for a long time, and more recent ones are rare and derived from the commonest haplotypes (Posada & Crandall 2001). Additionally, a root haplotype is expected to have a higher number of haplotype

© 2015 John Wiley & Sons Ltd

connections in the network, rather than being close to the tips. However, past extinction of haplotypes can obscure the inference, with younger haplotypes becoming the most prevalent, so accurately identifying the root haplotype is a challenging task. The fact that BPEC does not provide a single estimate of the haplotype network like Statistical Parsimony, but a finite (probability) distribution of haplotype trees—as well as the existence of the underlying migration model—allows this method to incorporate the uncertainty in the haplotype rooting. BPEC estimates the missing, extinct haplotype H11 as the most probable root of the haplotype network (Fig. 3b), although haplotypes H1, H2 and H8 are also associated with high posterior probabilities.

Have palaeo-islands acted as sources of genetic diversity within and towards other islands?

The theory of Pleistocene climate refugia (Hewitt 2000) states that historically environmentally stable areas can act as sources of genetic diversity exporting migrant alleles to other, disturbed regions (Gavin *et al.* 2014). Palaeo-islands could have played the same role in volcanic archipelagos, although in this case catastrophic geological events rather than climatic changes might be responsible for the observed patterns. BPEC provides support to this hypothesis, identifying the populations in the palaeo-islands of Tenerife as the source areas of ancestral migration events to other adjacent islands, such as from Teno to La Gomera or from Anaga to Gran Canaria (Fig. 3a, b; Table 2). A third dispersal event from west Tenerife/La Gomera to La Palma is supported by the BEAST tree (Fig. 4), and a fourth dispersal event from La Palma to El Hierro is inferred by the BPEC analysis (Fig. 3). Similar patterns with the central Canaries as centres of dispersal events within the archipelago have been described in other animal and plant studies (Francisco-Ortega *et al.* 2002; Gómez *et al.* 2003; Sanmartín *et al.* 2008; Mairal *et al.* 2015; Puppo *et al.* 2015). Moreover, a spatio-temporal pattern of colonization comparable to *Canarina*, showing Tenerife as the centre of dispersal events to adjacent islands in the last 1 Ma, can be found in the Canarian lineage of *Cistus* [0.33 (0.88–0.07) Ma; Guzmán & Vargas 2010], *Cheirolophus* (1 Ma; Vitales *et al.* 2014a), *Cistus monspeliensis* (0.93–0.20 Ma; Fernández-Mazuecos & Vargas 2011) and *Gallotia galloti* (0.8–0.9 Ma; Cox *et al.* 2010). This might be a consequence of the central geographical position of Tenerife within the archipelago—acting as a crossroad for dispersal events—but also of the concentration of plant genetic diversity in the palaeo-islands.

The case of Gran Canaria is especially interesting. It shares the same haplotype with Tenerife (H1, Figs 2 and 3). Low haplotypic diversity (Hn and Hd in Table 1) could be explained by a recent colonization after a catastrophic event. The island was subject to intense volcanic activity during the Holocene (24 eruption events; Rodríguez-Gonzalez *et al.* 2009), so extinction might explain its present low genetic diversity. Another possibility is related to the topography of Gran Canaria, where a network of ravines (locally known as 'barrancos') connecting at their summits could have facilitated gene flow among populations. Dispersal of *C. canariensis* seeds by *Gallotia* lizards (Rodríguez *et al.* 2008), probably using forest gaps and edges of roads as dispersal corridors (Delgado *et al.* 2007), might have helped to connect populations in the highly altered laurel forest of Gran Canaria.

In addition to dispersal events between-islands, the palaeo-islands of Tenerife might have acted as sources of genetic diversity within Tenerife, exporting migrant alleles to other geologically unstable, disturbed areas. Our demographic analyses indicated a recent population expansion in two populations of east Tenerife close to Anaga (TF-Br and TF-Bj). Although this result should be taken with caution (the EBSP analysis supported a constant effective population size; Fig. S2, Supporting information), these two populations exhibited also a single cpDNA haplotype, which agrees with the idea of a recent colonization. The areas where these populations are located (La Orotava Valley and Güimar Valley, respectively) have been subjected to catastrophic volcanic events (Fig. 1b). Some authors (Thorpe *et al.* 1996; Gübitz *et al.* 2000) have proposed the existence of a corridor of suitable habitat along the northern coastal fringe of Tenerife to explain migrations of the reptiles *G. galloti* and *T. delalandii* from Anaga to the west. These dispersal events could also explain the patterns found in our BPEC analyses, which suggest Anaga populations as sources of migration events to other populations in eastern Tenerife and Gran Canaria (Fig. 3; Table 2).

Ongoing genetic connectivity between the populations in the palaeo-islands of Tenerife and those from other areas is supported by the nuclear genome, which shows genetic admixture between west-Anaga, Teno and La Gomera (Fig. 5a). Admixture could be explained by the carrying of pollen by nectar-feeding birds between forest patches (e.g. ringing techniques have confirmed migration of the main pollinator (*Phylloscopus canariensis*) between Teno and Anaga; A. González, personal communication). The fact that this connectivity is to some extent lost in the cpDNA might be explained by the cpDNA not being transported via pollen or, alternatively, by the small size and haploid nature of the chloroplast genome, which imply shorter coalescent times and less time to fix novel mutations for chloroplast markers (Avise 2000; Jakob & Blattner

© 2015 John Wiley & Sons Ltd

2006). On the other hand, the widespread distribution of some cpDNA haplotypes across the archipelago (H1, H3 and H9, Fig. 2a) supports some gene flow among populations driven by seed dispersal. In the Canary Islands, birds have been cited as important vectors for the dispersal of fleshy fruits (Arevalo *et al.* 2007; Padilla *et al.* 2012), and the latter has been associated with frequent gene flow preventing speciation within widespread island endemics (García-Verdugo *et al.* 2014), such as in *Canarina*. This fits with what is known on the reproductive biology of *C. canariensis*, which is pollinated by generalist birds (Rodríguez-Rodríguez & Valido 2011), while its fleshy fruits are dispersed by vertebrates such as *Gallotia* lizards (Valido *et al.* 2003; Rodríguez *et al.* 2008). An additional factor to explain frequent dispersal between-islands are the eustatic sea level shifts during the Pleistocene, which might have decreased geographical distance between-islands (Rijsdijk *et al.* 2014).

Finally, in addition to exporting migrant alleles to other islands and disturbed areas, the palaeo-islands of Tenerife might have acted as cradles or sources or new genetic diversity. The higher haplotype and nucleotide diversity and higher Nei's gene diversity Hj exhibited by populations in these areas (Table 1) agree with their role as ancient refugia but also as sources of novel genetic diversity. Maximum topographic complexity is one of the main factors explaining species richness and high speciation rates within-islands (Whittaker *et al.* 2007). The rugged nature of the Tenerife palaeo-islands has likely promoted genetic differentiation within these massifs. For example, divergences found between populations in east and west forest ranges within Anaga have been explained by the existence of deep ravines and the volcanic arc of Taganana (Fig. 1b), acting as geographical barrier to gene flow (Macías-Hernández *et al.* 2013). These divergences were also detected in *C. canariensis*.

Conclusions

Traditionally, the distribution of genetic diversity within archipelagos is assumed to be structured around oceanic barriers, with between-island divergences expected to be larger than within-island differentiation. Here, we showed that within-island genetic patterns might be as strong as or stronger than those observed between-islands when they are associated with historical volcanic events. In *C. canariensis*, geographical patterns of genetic variation are structured around the palaeo-islands of Tenerife, with a minor secondary effect due to oceanic barriers. Carine & Schaefer (2010) argued that although relatively short oceanic distances separate the Canary Islands, they might be responsible for the high diversity levels found in the archipelago, acting as effective barriers to dispersal and promoting allopatric speciation. However, our results suggest that this hypothesis might not be valid for endemic species with widespread distributions across several islands (e.g. *Canarina*), for whom stretches of ocean are apparently less of a barrier than topographic relief within volcanic islands. Phylogeographical studies on other MIEs (multiple island endemics) are needed to confirm this hypothesis. The palaeo-islands of Tenerife have probably acted as both genetic refugia and sources of new diversity within- and between-islands. The preservation of genotypes that became extinct everywhere else and the topographic complexity of the palaeo-islands makes them potential 'phylogeographical hotspots' (Médail & Diadema 2009) and reservoirs of unique genetic diversity, whose conservation should be prioritized.

Acknowledgements

We are grateful to Mitchell Cruzan and four anonymous reviewers, whose comments helped to significantly improve our manuscript. We would like to thank Alberto Herrero, Emilio Cano and Fatima Durán for their laboratory assistance. Field work could not have been conducted without the cooperation of the staff at the Jardín Botánico Canario Viera y Clavijo (Gran Canaria); Jacinto Leralta and Ángel Fernández (La Gomera); Félix Manuel Medina from the Cabildo of La Palma, and the Cabildo of Tenerife and El Hierro are thanked for helping with sampling permits and accommodation during field expeditions. We are grateful to Gonzalo Nieto-Feliner for a critical review of the manuscript. We also thank Águedo Marrero, Alfredo Valido, Aurelio Martín, Carlos García-Verdugo, Isabel Gómez, Juli Caujapé-Castells, Lisa Pokorny, María Candelaria Rodríguez-Rodríguez, Marta Martínez, Moisés Soto, Nadir Álvarez, Nils Arrigo, Rubén Barone, Xavier Picó and Yulán Úbeda for all the fruitful discussions. This work was funded by Spanish Ministry of Science through projects CGL2009-1332-C03-01 and CGL2012-40129-C02-01). M. Alarcón was funded by the JAE-Doc program (CSIC/FSE). MM and VC were supported by MINECO FPI predoctoral fellowships (BES-2010-037261 and BES-2013-065389, respectively).

References

Ancochea E, Fuster J, Ibarrola E *et al.* (1990) Volcanic evolution of the island of Tenerife (Canary Islands) in the light of new K-Ar data. *Journal of Volcanology and Geothermal Research*, **44**, 231–249.

Ancochea E, Huertas MJ, Cantagrel JM *et al.* (1999) Evolution of the Cañadas edifice and its implications for the origin of the Cañadas Caldera (Tenerife, Canary Islands). *Journal of Volcanology and Geothermal Research*, **88**, 177–199.

Arevalo JR, Delgado JD, Fernández-Palacios JM (2007) Variation in fleshy fruit fall composition in an island laurel forest of the Canary Islands. *Acta Oecologica*, **32**, 152–160.

Avise JC (2000) *Phylogeography: The History and Formation of Species*. Harvard University Press, Cambridge, Massachusetts.

© 2015 John Wiley & Sons Ltd

Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV (2012) Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Molecular Biology and Evolution*, **29**, 2157–2167.

Báez M, Martín JL, Oromí P (2001) Diversidad taxonómica terrestre. En (Ed.) Fernández Palacios JM & Martín Esquivel JL, Naturaleza de las Islas Canarias. Ecología y Conservación.Tenerife: Ed. Turquesa, 65–76.

Bloomquist EW, Lemey P, Suchard MA (2010) Three roads diverged? Routes to phylogeographic inference. *Trends in Ecology & Evolution*, **25**, 626–632.

Bonin A, Bellemain E, Bronken Eidesen P, Pompanon F, Brochmann C, Taberlet P (2004) How to track and assess genotyping errors in population genetics studies. *Molecular Ecology*, **13**, 3261–3273.

Bottin L, Verhaegen D, Tassin J, Olivieri I, Vaillant A, Bouvet JM (2005) Genetic diversity and population structure of an insular tree, Santalum austrocaledonicum in New Caledonian archipelago. *Molecular Ecology*, **14**, 1979–1989.

Boulesteix T, Hildenbrand A, Gillot PY, Soler V (2012) Eruptive response of oceanic islands to giant landslides: new insights from the geomorphologic evolution of the Teide-Pico Viejo volcanic complex (Tenerife, Canary). *Geomorphology*, **138**, 61–73.

Boulesteix T, Hildenbrand A, Soler V, Quidelleur X, Gillot PY (2013) Coeval giant landslides in the Canary Islands: implications for global, regional and local triggers of giant flank collapses on oceanic volcanoes. *Journal of Volcanology and Geothermal Research*, **257**, 90–98.

Brown RP, Campos-Delgado R, Pestano J (2000) Mitochondrial DNA evolution and population history of the Tenerife skink *Chalcides viridanus*. *Molecular Ecology*, **9**, 1061–1067.

Brown RP, Hoskisson PA, Welton JH, Baez M (2006) Geological history and within-island diversity: a debris avalanche and the Tenerife lizard *Gallotia galloti*. *Molecular Ecology*, **15**, 3631–3640.

Cantagrel JM, Arnaud NO, Ancochea E, Fúster JM, Huertas MJ (1999) Repeated debris avalanches on Tenerife and genesis of Las Cañadas caldera wall (Canary Islands). *Geology*, **27**, 739–742.

Carine MA, Schaefer H (2010) The Azores diversity enigma: why are there so few Azorean endemic flowering plants and why are they so widespread? *Journal of Biogeography*, **37**, 77–89.

Carracedo J (1994) The Canary Islands: an example of structural control on the growth of large oceanic-island volcanoes. *Journal of Volcanology and Geothermal Research*, **60**, 225–241.

Carracedo JC (2014) The Teide volcano, Tenerife, Canary Islands. In: *Landscapes and Landforms of Spain*, pp, 257–272. Springer, Netherlands

Carracedo JC, Day S (2002) *Canary Islands*. Dunedin Academic Press Ltd, Lewiston, New York.

Carracedo JC, Day S, Guillou H, Rodríguez Badiola E, Canas JA, Pérez Torrado FJ (1998) Hotspot volcanism close to a passive continental margin: the Canary Islands. *Geological Magazine*, **135**, 591–604.

Carracedo JC, Badiola ER, Guillou H *et al.* (2007) Eruptive and structural history of Teide Volcano and rift zones of Tenerife, Canary Islands. *Geological Society of America Bulletin*, **119**, 1027–1051.

Carson HL, Templeton AR (1984) Genetic revolutions in relation to speciation phenomena: the founding of new populations. *Annual Review of Ecology and Systematics*, **15**, 97–131.

Carson HL, Lockwood JP, Craddock EM (1990) Extinction and recolonization of local populations on a growing shield volcano. *Proceedings of the National Academy of Sciences*, **87**, 7055–7057.

Caujapé-Castells J (2010) General G_{ST} and θ inflation due to biased intra-population sampling, and its consequences for the conservation of the Canarian Flora. *Conservation Genetics*, **11**, 709–720.

Clement M, Posada DCKA, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1659.

Contreras-Díaz HG, Moya O, Oromí P, Juan C (2007) Evolution and diversification of the forest and hypogean ground-beetle genus *Trechus* in the Canary Islands. *Molecular Phylogenetics and Evolution*, **42**, 687–699.

Cox SC, Carranza S, Brown RP (2010) Divergence times and colonization of the Canary Islands by *Gallotia* lizards. *Molecular Phylogenetics and Evolution*, **56**, 747–757.

Delgado JD, Arévalo JR, Fernández-Palacios JM (2007) Road edge effect on the abundance of the lizard *Gallotia galloti* (Sauria: Lacertidae) in two Canary Islands forests. *Biodiversity and Conservation*, **16**, 2949–2963.

Dimitrov D, Arnedo MA, Ribera C (2008) Colonization and diversification of the spider genus *Pholcus* Walckenaer, 1805 (Araneae, Pholcidae) in the Macaronesian archipelagos: evidence for long-term occupancy yet rapid recent speciation. *Molecular Phylogenetics and Evolution*, **48**, 596–614.

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**, 214.

Earl DA (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

Eckert CG, Samis KE, Lougheed SC (2008) Genetic variation across species' geographical ranges: the central–marginal hypothesis and beyond. *Molecular Ecology*, **17**, 1170–1188.

Emerson BC, Oromí P (2005) Diversification of the forest beetle genus *Tharpius* on the Canary Island, and the evolutionary origins of island endemics. *Evolution*, **59**, 586–598.

Emerson BC, Oromı́ P, Hewitt GM (1999) MtDNA phylogeography and recent intra-island diversification among Canary Island *Calathus* beetles. *Molecular Phylogenetics and Evolution*, **13**, 149–158.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47.

Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes*, **7**, 574–578.

Feliner GN (2014) Patterns and processes in plant phylogeography in the Mediterranean basin. A review. *Perspectives in Plant Ecology, Evolution and Systematics*, **16**, 265–278.

Fernández-Mazuecos M, Vargas P (2011) Genetically depauperate in the continent but rich in oceanic islands: *Cistus monspeliensis* (Cistaceae) in the Canary Islands. *PLoS One*, **6**, e17172.

Fernández-Palacios JM, de Nascimento L, Otto R *et al.* (2011) A reconstruction of Palaeo-Macaronesia, with particular refer-

© 2015 John Wiley & Sons Ltd

ence to the long-term biogeography of the Atlantic island laurel forests. *Journal of Biogeography*, **38**, 226–246.

Francisco-Ortega J, Santos-Guerra A, Kim SC, Crawford DJ (2000) Plant genetic diversity in the Canary Islands: a conservation perspective. *American Journal of Botany*, **87**, 909–919.

Francisco-Ortega J, Fuertes-Aguilar J, Kim SC, Santos-Guerra A, Crawford DJ, Jansen RK (2002) Phylogeny of the Macaronesian endemic *Crambe* section Dendrocrambe (Brassicaceae) based on internal transcribed spacer sequences of nuclear ribosomal DNA. *American Journal of Botany*, **89**, 1984–1990.

Fu YX (1996) New statistical tests of neutrality for DNA samples from a population. *Genetics*, **143**, 557–570.

Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. *Genetics*, **133**, 693–709.

García-Verdugo C, Forrest AD, Fay MF, Vargas P (2010) The relevance of gene flow in metapopulation dynamics of an oceanic island endemic, *Olea europaea* subsp. guanchica. *Evolution*, **64**, 3525–3536.

García-Verdugo C, Baldwin BG, Fay MF, Caujapé-Castells J (2014) Life history traits and patterns of diversification in oceanic archipelagos: a meta-analysis. *Botanical Journal of the Linnean Society*, **174**, 334–348.

García-Verdugo C, Sajeva M, La Mantia T, Harrouni C, Msanda F, Caujapé-Castells J (2015) Do island plant populations really have lower genetic variation than mainland populations? Effects of selection and distribution range on genetic diversity estimates. *Molecular Ecology*, **24**, 726–741.

Gavin DG, Fitzpatrick MC, Gugger PF *et al.* (2014) Climate refugia: joint inference from fossil records, species distribution models and phylogeography. *New Phytologist*, **204**, 37–54.

Gillespie R (2004) Community assembly through adaptive radiation in Hawaiian spiders. *Science*, **303**, 356–359.

Gillespie RG, Clague DA (2009) *Encyclopedia of Islands*. University of California Press, Berkeley, California.

Gillespie RG, Roderick GK (2002) Arthropods on islands: colonization, speciation, and conservation. *Annual Review of Entomology*, **47**, 595–632.

Gillespie RG, Roderick GK (2014) Evolution: geology and climate drive diversification. *Nature*, **509**, 297–298.

Gómez A, González-Martínez SC, Collada C, Climent J, Gil L (2003) Complex population genetic structure in the endemic Canary Island pine revealed using chloroplast microsatellite markers. *Theoretical and Applied Genetics*, **107**, 1123–1131.

Gübitz T, Thorpe RS, Malhotra A (2000) Phylogeography and natural selection in the Tenerife gecko *Tarentola delalandii*: testing historical and adaptive hypotheses. *Molecular Ecology*, **9**, 1213–1221.

Gübitz T, Thorpe RS, Malhotra A (2005) The dynamics of genetic and morphological variation on volcanic islands. *Proceedings of the Royal Society B: Biological Sciences*, **272**, 751–757.

Guillot G, Estoup A, Mortier F, Cosson JF (2005) A spatial statistical model for landscape genetics. *Genetics*, **170**, 1261–1280.

Guillou H, Carracedo JC, Paris R, Pérez Torrado FJ (2004) Implications for the early shield-stage evolution of Tenerife from K/Ar ages and magnetic stratigraphy. *Earth and Planetary Science Letters*, **222**, 599–614.

Guzmán B, Vargas P (2010) Unexpected synchronous differentiation in Mediterranean and Canarian *Cistus* (Cistaceae). *Perspectives in Plant Ecology, Evolution and Systematics*, **12**, 163–174.

Harpending H (1994) Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Human Biology*, **66**, 591–600.

Harter DE, Irl SD, Seo B *et al.* (2015) Impacts of global climate change on the floras of oceanic islands–Projections, implications and current knowledge. *Perspectives in Plant Ecology, Evolution and Systematics*, **17**, 160–183.

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.

Ho SY, Phillips MJ, Cooper A, Drummond AJ (2005) Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Molecular Biology and Evolution*, **22**, 1561–1568.

Hochkirch A, Goerzig Y (2009) Colonization and speciation on volcanic islands: phylogeography of the flightless grasshopper genus *Arminda* (Orthoptera, Acrididae) on the Canary Islands. *Systematic Entomology*, **34**, 188–197.

von Humboldt A (1814) *Voyage aux regions equinoxiales du nouveaux continent, fait en 1799–1804 par Al. de Humboldt et A. Bonpland*. Part 1. París.

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, **23**, 254–267.

Jakob SS, Blattner FR (2006) A chloroplast genealogy of *Hordeum* (Poaceae): long-term persisting haplotypes, incomplete lineage sorting, regional extinction, and the consequences for phylogenetic inference. *Molecular Biology and Evolution*, **23**, 1602–1612.

Jones KE, Reyes-Betancort JA, Hiscock SJ, Carine MA (2014) Allopatric diversification, multiple habitat shifts, and hybridization in the evolution of *Pericallis* (Asteraceae), a Macaronesian endemic genus. *American Journal of Botany*, **101**, 637–651.

Juan C, Ibrahim KM, Oromi P, Hewitt GM (1996) Mitochondrial DNA sequence variation and phylogeography of *Pimelia* darkling beetles on the island of Tenerife (Canary Islands). *Heredity*, **77**, 589–598.

Juan C, Emerson BC, Oromí P, Hewitt GM (2000) Colonization and diversification: towards a phylogeographic synthesis for the Canary Islands. *Trends in Ecology & Evolution*, **15**, 104–109.

Katoh K, Misawa K, Kuma KI, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, **30**, 3059–3066.

Kelchner SA (2000) The evolution of non-coding chloroplast DNA and its application in plant systematics. *Annals of the Missouri Botanical Garden*, **87**, 482–498.

Kim SC, McGowen MR, Lubinsky P, Barber JC, Mort ME, Santos-Guerra A (2008) Timing and tempo of early and successive adaptive radiations in Macaronesia. *PLoS One*, **3**, e2139.

Knowles LL (2008) Why does a method that fails continue to be used? *Evolution*, **62**, 2713–2717.

Kunkel G (1991) Flora y vegetación del Archipiélago Canario. Tratado florístico. 2ª parte. Dicotiledóneas. 312 pp. Edirca. Madrid.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.

© 2015 John Wiley & Sons Ltd

Lyell C (1855) *A Manual of Elementary Geology*. John Murray, London.

Lynch M, Milligan BG (1994) Analysis of population genetic structure with RAPD markers. *Molecular Ecology*, **3**, 91–99.

Macías-Hernández N, Bidegaray-Batista L, Emerson BC, Oromí P, Arnedo M (2013) The imprint of geologic history on within-island diversification of woodlouse-hunter spiders (Araneae, Dysderidae) in the Canary Islands. *Journal of Heredity*, **104**, 341–356.

Mairal M, Pokorny L, Aldasoro JJ, Alarcón M, Sanmartín I (2015) Ancient vicariance and climate-driven extinction continental-wide disjunctions in Africa: the case of the Rand Flora genus Canarina (Campanulaceae). *Molecular Ecology*, **24**, 1335–1354.

Manni F, Guerard E, Heyer E (2004) Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by using Monmonier's algorithm. *Human Biology*, **76**, 173–190.

Manolopoulou I, Emerson BC (2012) Phylogeographic ancestral inference using the coalescent model on haplotype trees. *Journal of Computational Biology*, **19**, 745–755.

Manolopoulou I, Legarreta L, Emerson BC, Brooks S, Tavaré S (2011) A Bayesian approach to phylogeographic clustering. *Interface Focus*, **1**, 909–921.

Mansion G, Parolly G, Crowl AA *et al.* (2012) How to handle speciose clades? Mass taxon-sampling as a strategy towards illuminating the natural history of *Campanula* (Campanuloideae). *PLoS One*, **7**, e50076.

Médail F, Diadema K (2009) Glacial refugia influence plant diversity patterns in the Mediterranean basin. *Journal of Biogeography*, **36**, 1333–1345.

Médail F, Quézel P (1997) Hot-spots analysis for conservation of plant biodiversity in the Mediterranean basin. *Annals of the Missouri Botanical Garden*, **84**, 112–127.

Moya Ó, Contreras-Díaz HG, Oromí P, Juan C (2004) Genetic structure, phylogeography and demography of two ground-beetle species endemic to the Tenerife laurel forest (Canary Islands). *Molecular Ecology*, **13**, 3153–3167.

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences*, **76**, 5269–5273.

Nylander J (2004) *MrModeltest v2. Program Distributed by the Author*. Evolutionary Biology Centre, Uppsala University 2, Uppsala.

Olesen JM, Alarcon M, Ehlers BK, Aldasoro JJ, Roquet C (2012) Pollination, biogeography and phylogeny of oceanic island bellflowers (Campanulaceae). *Perspectives in Plant Ecology, Evolution and Systematics*, **14**, 169–182.

Padilla DP, González-Castro A, Nogales M (2012) Significance and extent of secondary seed dispersal by predatory birds on oceanic islands: the case of the Canary archipelago. *Journal of Ecology*, **100**, 416–427.

Patiño J, Carine M, Mardulyn P *et al.* (2015) Approximate Bayesian computation reveals the crucial role of oceanic islands for the assembly of continental biodiversity. *Systematic Biology*, **64**, 579–589.

de Paz JP, Caujapé-Castells J (2013) A review of the allozyme data set for the Canarian endemic flora: causes of the high genetic diversity levels and implications for conservation. *Annals of Botany*, **111**, 1059–1073.

Pokorny L, Oliván G, Shaw AJ (2011) Phylogeographic patterns in two southern hemisphere species of Calyptrochaeta (Daltoniaceae, Bryophyta). *Systematic Botany*, **36**, 542–553.

Posada D, Crandall KA (2001) Intraspecific gene genealogies: trees grafting into networks. *Trends in Ecology & Evolution*, **16**, 37–45.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Puppo P, Curto M, Velo-Antón G, Pérez de Paz PL, Meimberg H (2014) The influence of geological history on diversification in insular species: genetic and morphological patterns of Micromeria Benth. (Lamiaceae) in Tenerife (Canary archipelago). *Journal of Biogeography*, **41**, 1871–1882.

Puppo P, Curto M, Gusmão-Guedes J *et al.* (2015) Molecular phylogenetics of *Micromeria* (Lamiaceae) in the Canary Islands, diversification and inter-island colonization patterns inferred from nuclear genes. *Molecular Phylogenetics and Evolution*, **89**, 160–170.

Rambaut A, Drummond A (2009) FigTree v1. 3.1. Computer program and documentation distributed by the author at http://tree.bio.ed.ac.uk/software.

Rambaut A, Drummond AJ, Suchard M (2007) *Tracer v1. 6*. http://tree.bio.ed.ac.uk/software/tracer/.

Reyes-Betancort JA, Santos-Guerra A, Guma IR, Humphries CJ, Carine MA (2008) Diversity, rarity and the evolution and conservation of the Canary Islands endemic flora. *Anales del Jardín Botánico de Madrid*, **65**, 25.

Rijsdijk KF, Hengl T, Norder SJ *et al.* (2014) Quantifying surface-area changes of volcanic islands driven by Pleistocene sea-level cycles: biogeographical implications for the Macaronesian archipelagos. *Journal of Biogeography*, **41**, 1242–1254.

Rodríguez A, Nogales M, Rumeu B, Rodríguez B (2008) Temporal and spatial variation in the diet of the endemic lizard *Gallotia galloti* in an insular Mediterranean scrubland. *Journal of Herpetology*, **42**, 213–222.

Rodríguez-Gonzalez A, Fernandez-Turiel JL, Perez-Torrado FJ *et al.* (2009) The Holocene volcanic history of Gran Canaria island: implications for volcanic hazards. *Journal of Quaternary Science*, **24**, 697–709.

Rodríguez-Rodríguez MC, Valido A (2011) Consequences of plant-pollinator and floral-herbivore interactions on the reproductive success of the Canary Islands endemic *Canarina canariensis* (Campanulaceae). *American Journal of Botany*, **98**, 1465–1474.

Rohlf F (1998) *NTSYS-pc Version 2.0. Numerical Taxonomy and Multivariate Analysis System*. Exeter Software, Setauket, New York.

Salomone N, Emerson BC, Hewitt GM, Bernini F (2002) Phylogenetic relationships among the Canary Island Steganacaridae (Acari, Oribatida) inferred from mitochondrial DNA sequence data. *Molecular Ecology*, **11**, 79–89.

Sanmartín I, Van Der Mark P, Ronquist F (2008) Inferring dispersal: a Bayesian approach to phylogeny-based island biogeography, with special reference to the Canary Islands. *Journal of Biogeography*, **35**, 428–449.

Santos-Guerra A (2001) Flora vascular nativa. In: *Naturaleza de las Islas Canarias. Ecología y Conservación* (eds Fernández-Palacios JM, Martín-Esquivel JM), pp. 185–198. Publicaciones Turquesa, Santa Cruz de Tenerife.

© 2015 John Wiley & Sons Ltd

Stuessy TF, Takayama K, López-Sepúlveda P, Crawford DJ (2014) Interpretation of patterns of genetic variation in endemic plant species of oceanic islands. *Botanical Journal of the Linnean Society*, **174**, 276–288.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.

Templeton AR, Crandall KA, Sing CF (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics*, **132**, 619–633.

Thorpe RS, Black H, Malhotra A (1996) Matrix correspondence tests on the DNA phylogeny of the Tenerife lacertid elucidate both historical causes and morphological adaptation. *Systematic Biology*, **45**, 335–343.

Trusty JL, Olmstead RG, Santos-Guerra A, Sá-Fontinha S, Francisco-Ortega J (2005) Molecular phylogenetics of the Macaronesian-endemic genus Bystropogon (Lamiaceae): palaeo-islands, ecological shifts and interisland colonizations. *Molecular Ecology*, **14**, 1177–1189.

Tzedakis PC, Emerson BC, Hewitt GM (2013) Cryptic or mystic? Glacial tree refugia in northern Europe. *Trends in Ecology & Evolution*, **28**, 696–704.

Valido A, Nogales M, Medina FM (2003) Fleshy fruits in the diet of Canarian lizards *Gallotia galloti* (Lacertidae) in a xeric habitat of the island of Tenerife. *Journal of Herpetology*, **37**, 741–747.

Vekemans X, Beauwens T, Lemaire M, Roldán-Ruiz I (2002) Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. *Molecular Ecology*, **11**, 139–151.

Vitales D, García-Fernández A, Pellicer J *et al.* (2014a) Key processes for *Cheirolophus* (Asteraceae) diversification on oceanic islands inferred from AFLP data. *PLoS One*, **9**, e113207.

Vitales D, Garnatje T, Pellicer J, Vallès J, Santos-Guerra A, Sanmartín I (2014b) The explosive radiation of *Cheirolophus* (Asteraceae, Cardueae) in Macaronesia. *BMC Evolutionary Biology*, **141**, 118.

Vos P, Hogers R, Bleeker M *et al.* (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*, **23**, 4407–4414.

Wang Q, Zhou SL, Hong DY (2013) Molecular phylogeny of the platycodonoid group (Campanulaceae s. str.) with special reference to the circumscription of *Codonopsis*. *Taxon*, **62**, 498–504.

Whitlock R, Hipperson H, Mannarelli M, Butlin RK, Burke T (2008) An objective, rapid and reproducible method for scoring AFLP peak-height data that minimizes genotyping error. *Molecular Ecology Resources*, **8**, 725–735.

Whittaker RJ, Ladle RJ, Araújo MB, Fernández-Palacios JM, Delgado J, Arévalo J (2007) The island immaturity–speciation pulse model of island evolution: an alternative to the "diversity begets diversity" model. *Ecography*, **30**, 321–327.

Zaczek K, Troll V, Cachao M *et al.* (2015) Nannofossils in 2011 El Hierro eruptive products reinstate plume model for Canary Islands. *Scientific Reports*, **5**, 1–5.

Zhivotovsky LA (1999) Estimating population structure in diploids with multilocus dominant DNA markers. *Molecular Ecology*, **8**, 907–913.

Zink RM, Barrowclough GF (2008) Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology*, **17**, 2107–2121.

M.M., M.A., J.J.A. and I.S. designed the study; M.M. carried out the field work, M.A. and J.J.A. contributed some samples; M.M. performed the research and analysed the data under the supervision of M.A. and I.S.; I.M., M.M., I.S. and V.C. performed the BPEC analyses; M.M. wrote the paper with help from I.S., M.A. and J.J.A., M.M. and I.S. revised the manuscript. All authors contributed with comments and approved the final version.

Data accessibility

DNA sequences: Genbank Accession nos KP797991–KP798432.

GenBank accessions, sampling locations and/or online-only appendices uploaded as online supplemental material.

NEXUS files for the concatenated cpDNA data set and original AFLPs Matrix of *Canarina canariensis*, Dryad doi:10.5061/dryad.j4103.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Plant species which are restricted and shared between two or more of the ancient static areas considered: the three paleo-islands of Tenerife and La Gomera.

Table S2 Sampling information and GenBank accession numbers for all of the taxa included in this study.

Table S3 Results from the sensitivity analyses to assess confidence in our age estimates.

Table S4 Haplotypes of *Canarina canariensis*, as determined by using the variable sites and indels found in the sixteen *C. canariensis* populations with markers *Pet*B-*Pet*D (937 bp), *Rpl*32-*trn*L (654 bp) and *trn*S-*trn*G (658 bp).

Table S5 G_{ST} and number of migrants (in parentheses) obtained from cloroplastic data in *Canarina canariensis*.

Table S6 Mean ages and 95% HPD confidence intervals from the BEAST analysis.

Table S7 Results from the DNA polymorphism, neutrality test and mismatch raggedness for *Canarina canariensis* haplotypes.

Table S8 Pairwise F_{ST} between populations of *Canarina canariensis* with nuclear data.

Table S9 Hierarchical analysis of molecular variance (AMOVA) for *C. canariensis* based on allelic variation at different levels: (a) amongst groups, (b) amongst populations within groups and (c) within populations.

© 2015 John Wiley & Sons Ltd

Fig. S1 Clusters and networks obtained with bayesian phylogeographic and environmental clustering (BPEC) with plastid DNA sequences.

Fig. S2 Mismatch distribution (left handside) and Extended Bayesian Skyline Plot analysis (right handside) for the different groups.

Fig. S3 The MCC tree, from the BEAST analysis, showing nodes with mean ages and the 95% HPD confidence intervals (values specified in Table S5).

Fig. S4 Principal coordinates analysis of AFLP data for individuals of *C. canariensis*.

Fig. S5 The estimated probability of the likelihood function according to the Evanno method for: (a) STRUCTURE analysis for *C. canariensis*; (b) STRUCTURE analysis using only Teneriffean populations.

© 2015 John Wiley & Sons Ltd

View publication stats

And now onto the next adventure, wherever that may lead.

.

.

.

.

.

.

.

.

.

.

.

.

.

.

.

.

.

.

.

.

.

.

.

.

.

The End.